A Novel Approach to Imaging using a Dual Field-of-View Sensor

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy
by
Mitul M. Mehta

September 2016

Abstract

Most modern aircraft, such as missile systems and unmanned aerial vehicles have limited size, weight, power and cost (SWaP-C) capability. As the defence budget for military forces such as the UK and US continue to shrink, the emphasis on SWaP-C continues to strengthen. Military forces require smart weapons capable of precision strike, with a priority on safety. System manufacturers understand these requirements and limitations, and in response, develop miniaturised systems and components and also aim to consolidate these, into a single miniaturised solution. The growth of remotely operated aircraft, offers an ever present need for better, cheaper imaging systems. In general, sensors and seekers tend to be the biggest contribution to the cost and weight of an aircraft. Often, multiple imaging systems are needed dependent on the operational requirements.

In this thesis, a novel dual field-of-view imaging system/seeker is proposed, which uses a single imaging sensor to superimpose both a wide field-of-view and a narrow field-of-view image of the same scene, co-boresighted. This allows multiple operational requirements to function simultaneously. The wide field-of-view allows for continuous monitoring and surveillance of an area, whilst the narrow field-of-view enables target detection, identification and tracking capabilities.

Secondly, this thesis proposes a novel image separation technique to facilitate the separation of the superimposed imagery, using only the geometric relationship between the two different fields-of-views. The separation technique is then extended to operate over sequential frames (i.e. video), and to function with fixed cameras that exhibit (un)desired camera motions, such as vibrations or "jitter". The image quality of the separation technique is broadly analysed over a range of images with varying image characteristics and properties. A novel image quality metric (IQM) was also proposed in this thesis, and was used to analyse the image quality of the recovered images, and its performance compared to already available IQMs.

Finally, the separation technique is enhanced to operate with motion cameras, which exhibit motions such as pan, tilt, zoom and rotate etc. The separation technique, in most cases, was found to provide image recovery, comparable to current image enhancement techniques, and moreover, found to be far more robust to errors in registration, compared to current techniques.

Initial hardware designs for the dual field-of-view imaging system, designed in conjunction with Prof. Andy Harvey from the University of Glasgow and Dr. James Babbington from Qioptiq Ltd., a lens design and manufacturing company, has also been presented.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Prof. Jason Ralph, for his continued encouragement and support throughout the duration of my PhD study, without his guidance, knowledge and teaching this thesis would not be possible. A special thank you to Dr. Elias Griffith, who has supported the progress of my research throughout the four years, with his indispensable software development expertise.

I would like to thank my beloved family, for their unwavering support, encouragement and belief to reach my academic and personal goals, and who have shared this journey with me, and will continue to. Words cannot express what you mean to me, without them I would not be where I am today. A heartfelt mention goes to my grandparents, who saw me begin this journey, but are no longer with us to see me reach this special moment.

I would also like to thank the city of Liverpool, for being a wonderful home to me and to all the amazing people I have met, and friends I have made throughout this journey, along with all the memories we have shared together. A big thank you to the wonderful Salsa family for the years of dancing and motivation, that kept me going. Thank you for making this unforgettable journey, so special to me.

Contents

List of Figures

Nomenclature

fps Frames Per Second

px Pixels

$dFoV$ Dual Field-of-View

FoV Field-of-View

$nFoV$ Narrow Field-of-View

$wFoV$ Wide Field-of-View

A2A Air-to-Air

BLIINDS BLind Image Integrity Notator using DCT Statistics

BRISQUE Blind Image Spatial Quality Evaluator

BTT Bank-to-turn

CAT Computerized Axial Tomography

CC Cross Correlation

CSF Contrast Sensitivity Function

CT Computed Tomography

DCT Discrete Cosine Transform

DIIVINE Identification-based Image Verity and INtegrity Evaluation

F-SSIM Fast Structural SIMilarity

FFT Fast Fourier Transform

FHD Full High Definition

FR-IQA Full Reference Image Quality Assessment

GBU Guided Bomb Unit

GIS Geographic Information System

GPS Global Positioning System

HVS Human Visual System

INS Inertial Navigation System

IQA Image Quality Assessment

IQIM Image Quality Information Metric

IQM Image Quality Metric

JFIF JPEG File Interchange Format

LASER Light Amplification by Stimulated Emission of Radiation

LBR Laser Beam Riding

LGB Laser Guided Bomb

LMS Least Median of Squares

MPEG Moving Picture Experts Group

MRI Magnetic Resonance Imaging

MS-SSIM Multi-Scale Structural SIMilarity

MSE Mean-Squared Error

NCC Normalised Cross Correlation

NR-IQA No Reference Image Quality Assessment

PC Phase Correlation

PN Proportional Navigation

PPN Pure Proportional Navigation

PSNR Peak Signal-to-Noise Ratio

QHD Quad High Definition

RANSAC RANdom SAmple Consensus

RR-IQA Reduced Reference Image Quality Assessment

S2A Surface-to-Air

SACLoS semi-active command-to-line-of-sight

SAL Semi-active Laser

SAU Safe Arm Unit

SIFT Scale Invariant Feature Transform

SR Super-resolution

SSDA Sequential Similarity Detection Algorithm

SSEQ Spatial-Spectral Entropy-based Quality

SSIM Structural SIMilarity

STT Skid-to-turn

SURF Speeded-Up Robust Features

SWaP-C Size, Weight, Power and Cost

TA Temporal Average

UAS(s) Unmanned Aerial System(s)

UAV(s) Unmanned Aerial Vehicle(s)

VIF Visual Information Fidelity

Chapter 1

Introduction

Projectiles and missiles have been used for millennia, dating back to the creation of slingshots and the bow and arrow. More recently, munitions have been the choice for conflict resolution. Initially, these munitions were incapable of being guided (unguided) after launch. Only during the Second World War was the technology developed to build guided missiles which had the capability to adapt their trajectory after launch. Guided missiles had been proposed much earlier but the technology and scientific capability were insufficiently developed at the time. In 1943, during the Second World War, the German forces engineered the Fritz-X radio controlled bomb, which was recorded as the first operational guided weapon to be successfully deployed [113]. Weapons have continued to be engineered ever since. In present day, the weapon capabilities have become highly sophisticated and complex (termed complex weapons). The majority of air launched weapons used in combat operations are guided. Complex weapons or precision guided munitions are developed with precision and safety in mind. Safety is an ever growing issue [77], from logistics and delivery platform, to troops on the ground, as well as collateral damage. Complex weapons are designed to be capable of performing a precision strike whilst minimising collateral damage. An example of this improvement in safety can be empirically shown by taking the Paveway guided bomb unit (GBU), as an example. In the early developments, the Paveway III (also termed GBU-24/B) was a 2000lb warhead, laser guided bomb (LGB) brought into service in 1983 with many variants [206]. Later in 2008, Raytheon developed the Paveway IV, brought into service in 2008, containing a 500lb warhead. It is a dual-mode global positioning system/inertial navigation system (GPS/INS) and LGB munition [91], improving accuracy and hence, a smaller warhead is appropriate to minimise collateral damage. In 2008 the RAF and FAA deemed necessary to develop a precision guided munition with minimal collateral damage, hence the Dual Mode Brimstone was developed [60] containing a high frequency millimetre-wave (mmW) radar homing with INS, identical to the original Brimstone with added semi-active laser (SAL) guidance capability for higher accuracy, both containing a 6.2kg warhead to improve safety. An

improved version of the Brimstone missile, the Brimstone 2 [11], is currently under production, equipped with a new airframe, improved millimetre-wave radar with semi-active laser dual mode seeker capability, and an insensitive munition (IM) rocket motor and warhead, designed for higher precision, and higher safety during handling and logistics but also to minimise collateral damage during service. Guided weapons can be broken down into 6 main subsystems [71, 89, 178]:

The **airframe** includes the main body/fuselage, wings, canards, tails, fins, nose and control surfaces, all designed for structural and aerodynamic efficiency.

The **seeker/sensor** can vary in configuration from active, semi-active and passive homing seekers, also between imaging (e.g. visible, infrared, UV) to non-imaging (e.g. radar, laser) and pure navigation.

The **guidance & control** subsystem includes ***actuators*** (e.g. hydraulic, pneumatic, electro-mechanical or piezoelectric) used to deflect the control surfaces (e.g. canards, wings, fins, tails). The ***guidance laws***, that calculate the accelerations required to steer the airframe towards its point of intercept, and the ***control system*** which calculate the deflections required to produce the required accelerations, also to apply the deflections, monitor and provide feedback to the required subsystems of the airframe's trajectory.

The **payload**, is dependent on the mission objectives. In military applications, this tends to be the warhead. Different types of explosive warheads range from shaped charged, blast or fragmentation type.

Fusing is the device which initiates the detonation of the munition. Most devices will incorporate a safe arm unit (SAU), which prevents the munition from arming during shipping, handling and storage [95]. The fuse can be activated through several options such as: ***remote detonation*** - from a remote location; ***proximity fuse*** – detonates when the fuse is within a predetermined distance to the target; ***time fuse*** – detonates after a set period of time has elapsed; ***impact fuse*** – detonates upon impact, typically when physically striking the target.

Propulsion can range from jet engines (e.g. turbojet, ramjet, pulsejet) to solid (or liquid) propellant engines.

Of the main subsystems, the seeker/sensor is the most expensive and often most complicated [88, 89]. The seeker/sensor aims to provide ***situational awareness*** for surveillance, target monitoring, collateral damage investigation. This allows the operator to monitor and assess the area of interest. ***Target identification*** allows detection, recognition and identification of target. ***Target tracking*** provides the ability to follow the targets movement e.g. moving vehicle, and ***countermeasure sensitivity***, i.e. ability to detect and avoid countermeasures deployed by the target.

Seekers and sensors are continuously being revised to provide better functionality for longer range, narrow field-of-view, target identification and tracking, alongside wider

field-of-view operation for situational awareness and surveillance purposes. Often these operational requirements can be in opposition to each other, and in certain cases, to fulfil the operational requirements, multiple seekers/imagers are required; and operate in a coordinated way such that the wide field-of-view ($wFoV$) imager instructs the narrow field-of-view ($nFoV$) imager to interrogate a possible target, resulting in higher sensor integration, complexity and thus higher costs.

This thesis considers a new type of seeker/imager that aims to provide both wide field-of-view situational awareness and narrow field-of-view target identification and tracking capabilities simultaneously using a single seeker/imager. The proposed imaging sensor operates by superimposing the co-boresighted wide field-of-view ($wFoV$) and the narrow field-of-view ($nFoV$) imagery, of the same scene, into a single superimposed image frame. This thesis will then propose and analyse a novel computational imaging technique using the geometric relationship between the two fields-of-view in order to separate the superimposed frame, furthermore the technique will be extended to function over multiple frames (i.e. video) and to analyse the performance qualitatively and quantitatively. A new image quality metric is also presented, and the performance to the metric is analysed against, already available image metrics.

Although missile seekers/sensors are the main focus of this thesis, the proposed imaging system is highly applicable for other autonomous air platforms such as unmanned aerial vehicles (UAVs). In general, aircraft aim to carry larger payloads, and to reduce size, weight, power and cost (SWaP-C) which is generally achieved by the miniaturisation of current airborne components and/or the consolidation of components into a single solution [59].

1.1 Structure of Thesis

This thesis will be divided into the following chapters:

- Chapter 2 - Simulation Model, describes the resources and theory required to generate a 6DoF simulation model, in order to study the trajectory data of the airframe and determine the types of motions an imaging system may undergo, during operational usage.

- Chapter 3 - Guidance Systems & Seekers, reviews the current systems available for the guidance of an aircraft, and the various types of seekers that enable target interception. This will allow to integrate a modern guidance and seeker system into the simulation model, and proposes the novel dual field-of-view imaging system.

- Chapter 4 - Image Separation Technique, provides the theory required to separate the superimposed images, produced by the dual field-of-view imager and assess

the quality of the recovered images.

- Chapter 5 - Image Registration, presents the precursory knowledge required to extend the separation technique to operate over sequential frames (i.e. video), and the best registration method for our application is chosen.

- Chapter 6 - Temporal Averaging Geometric Separation Technique, presents a multi-frame image separation technique to be used with the novel imaging system. The chapter also demonstrates and analyses the quality of the recovered images.

- Chapter 7 - Image Quality Analysis. This chapter provides a brief overview of the different types of image quality assessment techniques available. The appropriate image quality metrics, alongside a new image quality metric also proposed in this chapter, are chosen to provide a broad analysis of the separation technique's performance over a collection of images. A new and improved separation technique is proposed to extend the functionality of the imaging system and an initial hardware design for the imaging system is also presented.

- Chapter 8 - Summary & Future Work, concludes the thesis, summarising the research carried out and the contributions made towards a new imaging system architecture. Recommendations for future work are also provided.

1.2 Original Contributions

This thesis, in conjunction with the published conference papers [179, 180], proposes a novel dual field-of-view ($dFoV$) imager/sensor system, which provides both wide field-of-view ($wFoV$) and narrow field-of-view ($nFoV$) imagery, of the same scene simultaneously. Conventional systems would require two individual camera systems for the $wFoV$ and $nFoV$ imagery, which operate in a coordinated way such that the $wFoV$ instructs the $nFoV$ imager to interrogate possible objects of interest. Both field-of-view cameras are necessary for different operational functions. The situational information obtained from the wide field-of-view allows for continuous monitoring of the general area, whilst the narrow field-of-view allows for closer interrogation and identification of objects of interest. The proposed $dFoV$ imaging sensor, will superimpose both field-of-view images into a single superimposed image frame. This thesis further proposes a technique, that has been developed to separate the superimposed images, furthermore, the research is extended to function over multiple frames (i.e. videos) and the performance of the technique is qualitatively and quantitatively evaluated. A new image quality assessment metric is also introduced, and its performance is analysed against current image quality metrics available.

The results of this work are focused on missile seekers but applicable to most UAV applications, most noticeably in military applications. In general however, all UAVs

which require an on-board imaging system can benefit from such a system.

1.3 Publications

1. M. Mehta, E. J. Griffith, S. Maskell and J. F. Ralph. Geometric separation of superimposed images. In IEEE 19th International Conference on Information Fusion (FUSION), Heidelberg, Germany, 2016. IEEE. [180].

2. M. Mehta, E. J. Griffith, and J. F. Ralph. Geometric separation of superimposed images with varying fields-of-view. In IEEE 17th International Conference on Information Fusion (FUSION), Salamanca, Spain, 2014. IEEE. [179].

3. C. Mishra, M. Mehta, E. J. Griffith, and J. F. Ralph. Doing the right thing: Collision avoidance for autonomous air vehicles. In Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, pg. 2581-2586, Oct 2013. IEEE. [188].

Chapter 2

Simulation Model

2.1 Introduction

Rapid development in modern computing and processing power provides useful benefits to engineers. Engineers can utilise the computing power to create simulation models where they are able to develop ideas and create a safe environment by which their ideas can be tested and/or proto-typed. Simulations can also provide a safe digital environment for training purposes or for optimising systems by testing different parameters and operational ranges without physical testing.

The work presented in this thesis focuses on imaging systems fitted onboard complex weapon systems and/or unmanned aerial vehicles (UAVs). The dynamical model that simulates such an aircraft would be beneficial in order to generate a realistic environment that can provide realistic trajectory data and imitate an aircraft, as well as the camera motions (i.e. panning, tilting, zooming) that can externally influence the performance of the imaging system in operational use.

In this chapter, the fundamental theory required to create a simulation model of a complex weapon is discussed. Topics include aerodynamics, propulsion, guidance and control. Furthermore this chapter aims to describe how such a simulation model is implemented. Additionally, it aims to present and analyse the performance of the simulation model that was created for the purpose of the work.

2.2 Six Degrees of Freedom Model

The six degrees of freedom model (6DoF) describes an arbitrary rigid body, which is free to move in the three translational degrees of freedom (vertical (Z_B), lateral (Y_B) and longitudinal (X_B)) and the three rotational degrees of freedom (pitch, roll and yaw) in three dimensional space [194] [299], as illustrated in Figure 2.2. In this case we focus on aerodynamic bodies such as aircraft, missiles, or bombs.

The dynamics are given by Newton's equations of motion, for a rigid body (in 6DoF)

as [296]:

$$\sum \vec{F} = \frac{d}{dt}(m\vec{V}) \tag{2.1}$$

where $\sum \vec{F}$ is the summation of the forces impinged on the rigid body.

$$\sum \vec{M} = \frac{d\vec{H}}{dt} \tag{2.2}$$

where $\sum \vec{M}$ is the summation of the moments impinged on the rigid body. or explicitly expressed (in body axes) as [299]:

$$F_X = m(\dot{u} + wQ - vR) \tag{2.3}$$

$$F_Y = m(\dot{v} + uR - wP) \tag{2.4}$$

$$F_Z = m(\dot{w} + vP - uQ) \tag{2.5}$$

$$M_X = \dot{P}I_{XX} - \dot{R}I_{XZ} + QR(I_{ZZ} - I_{YY}) - PQI_{XZ} \tag{2.6}$$

$$M_Y = \dot{Q}I_{YY} - PR(I_{XX} - I_{ZZ}) + (P^2 - R^2)I_{XZ} \tag{2.7}$$

$$M_Z = \dot{R}I_{ZZ} - \dot{P}I_{XZ} + PQ(I_{YY} - I_{XX}) - QRI_{XZ} \tag{2.8}$$

where,

$\vec{F} = (F_X, F_Y, F_Z)$ is the external force	$\vec{H}$ is the angular momentum vector
$\vec{\omega} = (P, Q, R)$ is the angular velocity vector	$\vec{V} = (u, v, w)$ is the velocity vector
$\vec{M} = (M_X, M_Y, M_Z)$ is the moment vector	$\vec{I}$ is the inertia matrix

The explicit 6DoF equations [2.3] - [2.8] are highly non-linear and coupled. A simple, yet effective approach is to linearise the problem using linear aerodynamic theory and Taylor series expansion, as described further in section 2.3.

Newton's equations of motion, are used to formulate the dynamics of a rigid body for 6DoF analysis. The equations are all based in the body axes i.e. the translations and rotations are described but it does not account for the global position of the object. Also, the motions described in the body frame of reference, are all relative motions of the body and need to be referred to an external frame of reference, in order to determine the actual motion and orientation. Therefore, earth axes are also used to calculate the actual/global motion of the airframe/rigid body with respect to the earths surface. It is important to know the position of an aircraft with respect to the earth axes for situational awareness purposes. From a designer's perspective it is necessary to work in both axes systems, in order to determine the performance characteristics of the aircraft.

2.3 Aerodynamics

Aerodynamics is the field of determining the motion of air and its interaction with solid objects. Aerodynamics is a sub-branch of fluid mechanics, focusing only on the interaction of objects with air. The aerodynamic equations that describe the interaction of air with solid objects (in our case, an aircraft body), are highly non-linear. In general, designers tend to linearise these governing equations using Taylor series expansion theorem and the small angle approximation, which limits the control surface deflections to $10° - 12°$ because it simplifies the control systems [7]. In aerodynamic theory there are six aerodynamic forces and moments that can be described by their non-dimensional coefficients [8]. The aerodynamic coefficients (Equations (2.9)-(2.14)) determine the aerodynamic characteristics of an aircraft and are primarily influenced by factors which affect the dynamic pressure i.e. aerodynamics angles (control surface deflections), Mach number, Reynolds number and propulsion system [247].

Axial force (drag) co-efficient, C_X:

$$C_X = C_{X_0} + C_{X_\alpha}\alpha + C_{X_\beta}\beta + C_{X_{\delta P}}\delta P + C_{X_{\delta q}}\delta q + C_{X_{\delta r}}\delta r + C_{X_{\delta X}}\delta X + C_{X_{\alpha^2}}\alpha^2 + C_{X_{\beta^2}}\beta^2 + C_{X_{\delta P^2}}\delta P^2 + C_{X_{\delta q^2}}\delta q^2 + C_{X_{\delta r^2}}\delta r^2 + C_{X_{\delta X^2}}\delta X^2 + \frac{d}{2\nu}C_{X_p}P + C_{X_q}Q + C_{X_r}R \tag{2.9}$$

Lateral force (side) co-efficient, C_Y:

$$C_Y = C_{Y_0} + C_{Y_\alpha}\alpha + C_{Y_\beta}\beta + C_{Y_{\delta P}}\delta P + C_{Y_{\delta q}}\delta q + C_{Y_{\delta r}}\delta r + C_{Y_{\delta X}}\delta X + C_{Y_{\alpha^2}}\alpha^2 + C_{Y_{\beta^2}}\beta^2 + C_{Y_{\delta P^2}}\delta P^2 + C_{Y_{\delta q^2}}\delta q^2 + C_{Y_{\delta r^2}}\delta r^2 + C_{Y_{\delta X^2}}\delta X^2 + \frac{d}{2\nu}C_{Y_p}P + C_{Y_q}Q + C_{Y_r}R \tag{2.10}$$

Normal force (lift) co-efficient, C_Z:

$$C_Z = C_{Z_0} + C_{Z_\alpha}\alpha + C_{Z_\beta}\beta + C_{Z_{\delta P}}\delta P + C_{Z_{\delta q}}\delta q + C_{Z_{\delta r}}\delta r + C_{Z_{\delta X}}\delta X + C_{Z_{\alpha^2}}\alpha^2 + C_{Z_{\beta^2}}\beta^2 + C_{Z_{\delta P^2}}\delta P^2 + C_{Z_{\delta q^2}}\delta q^2 + C_{Z_{\delta r^2}}\delta r^2 + C_{Z_{\delta X^2}}\delta X^2 + \frac{d}{2\nu}C_{Z_p}P + C_{Z_q}Q + C_{Z_r}R \tag{2.11}$$

Roll moment co-efficient, C_l:

$$C_l = C_{l_0} + C_{l_\alpha}\alpha + C_{l_\beta}\beta + C_{l_{\delta P}}\delta P + C_{l_{\delta q}}\delta q + C_{l_{\delta r}}\delta r + C_{l_{\delta X}}\delta X + C_{l_{\alpha^2}}\alpha^2 + C_{l_{\beta^2}}\beta^2 + C_{l_{\delta P^2}}\delta P^2 + C_{l_{\delta q^2}}\delta q^2 + C_{l_{\delta r^2}}\delta r^2 + C_{l_{\delta X^2}}\delta X^2 + \frac{d}{2\nu}C_{l_p}P + C_{l_q}Q + C_{l_r}R \tag{2.12}$$

Pitch moment co-efficient, C_m:

$$C_m = C_{m_0} + C_{m_\alpha}\alpha + C_{m_\beta}\beta + C_{m_{\delta P}}\delta P + C_{m_{\delta q}}\delta q + C_{m_{\delta r}}\delta r + C_{m_{\delta X}}\delta X + C_{m_{\alpha^2}}\alpha^2 + C_{m_{\beta^2}}\beta^2 + C_{m_{\delta P^2}}\delta P^2 + C_{m_{\delta q^2}}\delta q^2 + C_{m_{\delta r^2}}\delta r^2 + C_{m_{\delta X^2}}\delta X^2 + \frac{d}{2\nu}C_{m_p}P + C_{m_q}Q + C_{m_r}R \tag{2.13}$$

Yaw moment co-efficient, C_n:

$$C_n = C_{n_0} + C_{n_\alpha}\alpha + C_{n_\beta}\beta + C_{n_{\delta P}}\delta P + C_{n_{\delta q}}\delta q + C_{n_{\delta r}}\delta r + C_{n_{\delta X}}\delta X + C_{n_{\alpha^2}}\alpha^2 + C_{n_{\beta^2}}\beta^2 + C_{n_{\delta P^2}}\delta P^2 + C_{n_{\delta q^2}}\delta q^2 + C_{n_{\delta r^2}}\delta r^2 + C_{n_{\delta X^2}}\delta X^2 + \frac{d}{2\nu}C_{n_p}P + C_{n_q}Q + C_{n_r}R \tag{2.14}$$

where the subscripts indicate that the coefficients correspond to derivatives:

α = Angle of attack (rad)

β = Angle of side-slip (rad)

δp = Control surface deflection used to roll the airframe (rad)

δq = Control surface deflection used to pitch the airframe (rad)

δr = Control surface deflection used to yaw the airframe (rad)

δx = Control surface deflection used to slow the airframe (rad)

$\alpha^2, \beta^2, \delta p^2, ...$ = The squared values of the terms above, where the sign of the number is maintained.

$\dot{\alpha}$ = Rate of change of $\alpha(rad/s)$

$\dot{\beta}$ = Rate of change of side-slip (rad/s)

P = Roll rate

Q = Pitch rate

R = Yaw rate.

These coefficients are generated by taking into account the various shapes of the aerodynamic body and control surfaces and calculating the summation of each individual components, in order to determine the effect on the overall behaviour of the airframe (i.e. position, velocity, acceleration).

The respective forces can be calculated using [27]:

$$F_X = -mg\sin(\theta) + T - \frac{1}{2}\rho\nu^2 s C_X \tag{2.15}$$

$$F_Y = mg\cos(\theta)\sin(\phi) - \frac{1}{2}\rho\nu^2 s C_Y \tag{2.16}$$

$$F_Z = mg\cos(\theta)\cos(\phi) - \frac{1}{2}\rho\nu^2 s C_Z \tag{2.17}$$

where,

m = Mass
T = Thrust generated by aircraft
u = Velocity along X axis (body)
v = Velocity along Y axis (body)
w = Velocity along Z axis (body)
g = Gravitational constant ($\approx 9.81 m/s^2$)
s = Aerodynamic reference area (m^2)
P = Pitch rate X axis (body)
Q = Pitch rate Y axis (body)
R = Pitch rate Z axis (body)

The accelerations can then be calculated as:

$$A_X = -wQ + vR + (F_X)/m \tag{2.18}$$

$$A_Y = -uR + wP + (F_Y)/m \tag{2.19}$$

$$A_Z = -vP + uQ + (F_Z)/m \tag{2.20}$$

Using a similar procedure, the aircraft moments can be calculated, using [27]:

$$L = \frac{1}{2}\rho V^2 s C_l l \tag{2.21}$$

$$M = \frac{1}{2}\rho V^2 s C_m l \tag{2.22}$$

$$N = \frac{1}{2}\rho V^2 s C_n l \tag{2.23}$$

where,

ρ = Air density (kg/m^3)	m = Mass of the weapon (kg)
V = Aircraft velocity (m/s)	g = Gravity constant $(= 9.81 m/s^2)$
s = Reference Area (m^2)	α = Angle of Attack $(rads)$
β = Angle of side-slip $(rads)$	l = Length of moment arm (m)

Once the moments have been determined, the change in angular velocities can be calculated using the airframe's current rotation rates and its moments of inertia (for an axially symmetric airframe):

$$\frac{\partial P}{\partial t} = (-RQ(I_{ZZ} - I_{YY}) + L)/I_{XX} \tag{2.24}$$

$$\frac{\partial Q}{\partial t} = (-PR(I_{XX} - I_{ZZ}) + M)/I_{YY} \tag{2.25}$$

$$\frac{\partial R}{\partial t} = (-PQ(I_{YY} - I_{XX}) + N)/I_{ZZ} \tag{2.26}$$

I_{XX} = moment of inertia $(_{XX})$
I_{YY} = moment of inertia $(_{YY})$
I_{ZZ} = moment of inertia $(_{ZZ})$

The moments of inertia, can be calculated using structural analysis based on cantilever beam theory [196]:

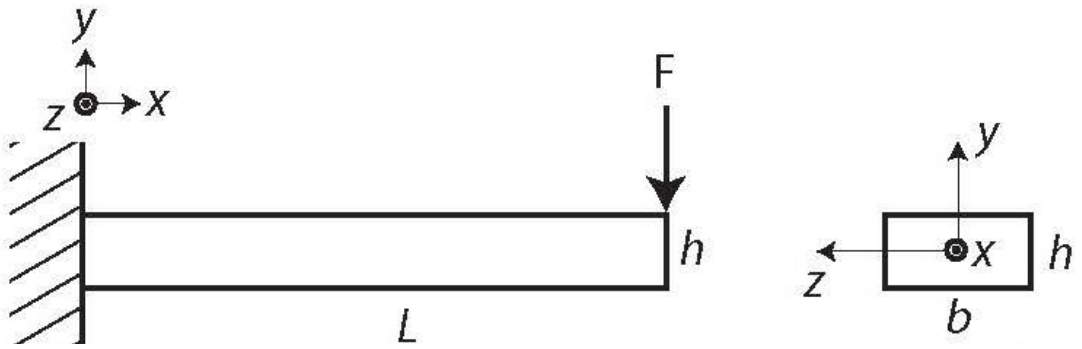


Figure 2.1: Illustration of cantilever setup for determination of the moments of inertia of an airframe. Image obtained from [127]

The moment of inertia along each airframe axis, can be calculated using:

$$I_{XX/YY/ZZ} = \frac{bh^3}{12} \tag{2.27}$$

where, b is the width/breadth of the structure and h is the height of the structure. The structure can be either the wings, fuselage or in fact any of the control surfaces.

To solve the problem numerically, the simulation model needs to discretise the problem to calculate the changes in the model over time. This allows to determine the rate of change of angular velocity, as follows:

$$\Delta P = \frac{\partial P}{\partial t}\Delta T \tag{2.28}$$

$$\Delta Q = \frac{\partial Q}{\partial t}\Delta T \tag{2.29}$$

$$\Delta R = \frac{\partial R}{\partial t}\Delta T \tag{2.30}$$

where, ΔT is the discrete model time-step, and $\partial/\partial t$ is the partial differential with respect to time. From this, the Euler angle rate can be determined by [241] (pp.334-335),

$$\frac{\Delta\theta}{\Delta t} = (Q + \Delta Q)\cos\phi - (R + \Delta R)\sin\phi \tag{2.31}$$

$$\frac{\Delta\psi}{\Delta t} = \frac{(Q + \Delta Q)\sin\phi + (R + \Delta R)\cos\phi}{\cos\theta} \tag{2.32}$$

$$\frac{\Delta\phi}{\Delta t} = (P + \Delta P) + \left(\frac{\Delta\psi}{\Delta t}\right)\sin\theta \tag{2.33}$$

where θ, ψ and ϕ are the heading (yaw), pitch and roll angles, respectively, in radians. Using a similar procedure as before, the *change* in Euler angles can be calculated using the model's time step:

$$\Delta\theta = \frac{\Delta\theta}{\Delta t}\Delta T \tag{2.34}$$

$$\Delta\psi = \frac{\Delta\psi}{\Delta t}\Delta T \tag{2.35}$$

$$\Delta\phi = \frac{\Delta\phi}{\Delta t}\Delta T \tag{2.36}$$

The equations expressed above describe the *changes* in the state of the aircraft. These changes are then summed to the model's existing states to calculate the up-to-date values, at the current time-step until the simulation is terminated.

2.4 Axes Transformation

As expressed in Section 2.2, the body axes allow to determine the changes in the force and motion impinged on the aircraft's body, but the earth coordinate reference system is required to determine its actual position, velocities and accelerations with respect to the earth's surface to fully describe an aircraft's motion.

The 6DoF model describes the aircraft's translational (u,v,w) and angular (P,Q,R) velocities which are in the airframe's body axes (see Figure 2.2, but in order to determine the actual position (X, Y, Z) and orientation (ψ (Yaw), θ (Pitch) and ϕ (Roll)) the earth coordinate reference system is needed (See Figure 2.3).

Following [107], transforming between sets of axes requires that a particular order of rotations are applied. Applying the rotations in a different order will yield a different solution. By convention, the rotation sequence commonly used is Z-Y-X (also known as 3-2-1). Hence, first the rotation in the Z axis (yaw) is applied, followed by the rotation in the Y axis (pitch) and finally the rotation in the X axis (roll) is applied, to convert from body axes to earth axes [38].

The earth axes are a positive right handed system, which is used to determine the true position of the airframe. The X and Y axes lie in the horizontal plane (X_E, Y_E) with the Z axis pointing vertically down in the direction of gravity (Z_E). The earth axes system is illustrated in Figure 2.3.

The three relative rotation matrices are given by:

$$3)\begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad 2)\begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \qquad 1)\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix}$$

The body coordinate system to earth coordinate system transformation matrix, T_E^B is expressed as the multiplication of the matrices:

$$T_E^B = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \tag{2.37}$$

Performing the inverse of this set of transforms will allow the conversion from earth axes to body axes , i.e. T_B^E.

$$T_B^E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.38}$$

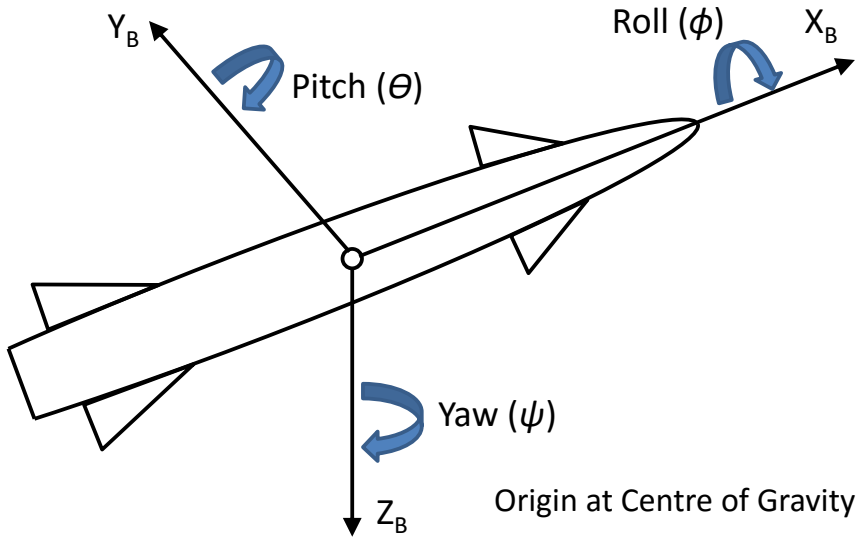


Figure 2.2: Translational and rotational degrees of freedom of airframe (i.e. body axes)

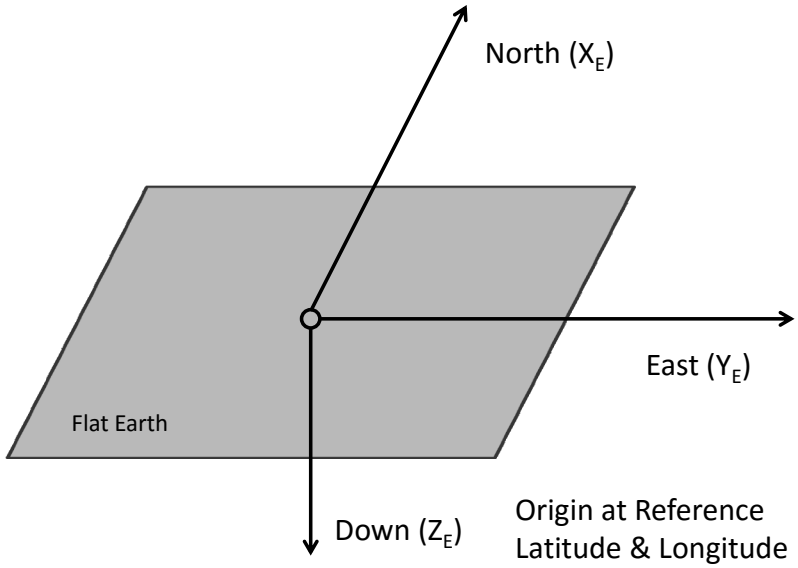


Figure 2.3: North-East-Down (NED) reference coordinate system (i.e. earth axes)

2.5 Propulsion

Most short range missiles commonly contain a two stage solid rocket fuel motor [71]. The two stages are initiated for different stages of the missile's flight. The first stage of the rocket motor, is known as the boost phase and is activated in the launch phase of the missiles flight, usually at this point no guidance is active and the main objective is to get the missile into the air at the optimal altitude and airspeed. The second stage of the rocket motor, known as the sustain phase, is initiated in the cruise phase of the missile where the guidance is activated and the airframe is steered towards the general position of the target. The fuel's burn rate is predetermined, but aims to maintain the cruise velocity, and to maximise the range of the missile, whilst being steered towards the target. The final stage, usually termed the terminal phase is when the airframe is directly homing onto the target, usually no propulsion is needed, as the airframe will be propelled by gravity towards the intended target [71].

The thrust of the solid rocket fuel can be characterised as follows:

$$T = I_{sp}\frac{dm}{dt}g \tag{2.39}$$

where I_{sp} is the specific impulse (seconds) and $\frac{dm}{dt}$ is the mass burn rate. The specific impulse is a characteristic parameter of the rocket fuel and it a way to representing the performance of the fuel. It describes the change in momentum produced per unit of propellant burnt [109, 197].

It is important to note that as a common rule generated from existing missiles as mentioned in [88] is that 75% of the fuel is used in the first stage (boost phase) and 25% of the fuel mass is used to power the second stage of propulsion (sustain phase).

The burn rate of the solid rocket fuel of weapon systems can be manipulated by changing the volume and cross-section profile of the solid fuel, as illustrated in Figure 2.4 below.

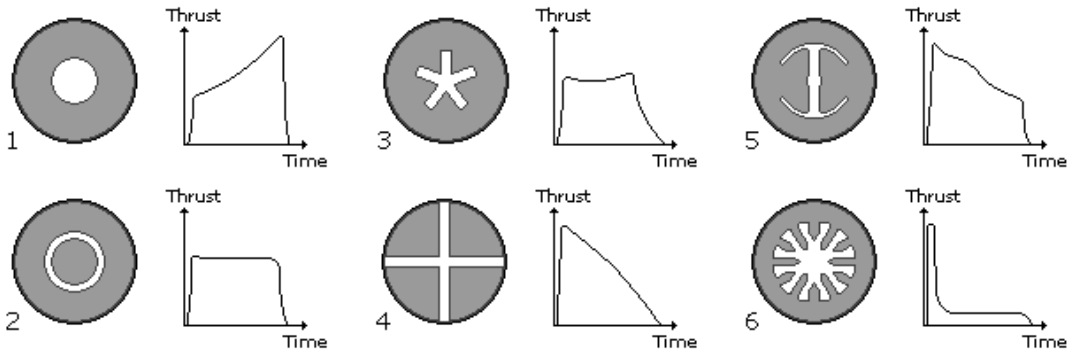


Figure 2.4: Solid rocket propellant cross-section & thrust profile [32]

2.6 Static Stability

Before incorporating the guidance laws it is important to assess the longitudinal stability of the simulation model. Having a stable airframe is important in terms of guidance, as if the airframe is unstable, the guidance laws will be generating very large lateral acceleration (latax) which will in turn increase the drag on the airframe and hence reduce the operational range of the missile.

The condition for a statically stable airframe is that the rate of change of pitching moment with respect to α ($\frac{dM}{d\alpha}$ or M_α) should be negative, but more essentially it should be a small negative value, otherwise if the value is too negatively large the airframe would become too stable and hinder the manoeuvrability of the missile [225]. To understand this condition the figure below (Figure 2.5) shows the stable and unstable conditions. The equivalent can be evaluated for the lateral (side-slip) stability with respect to β (i.e. $\frac{dM}{d\beta}$ or M_β).

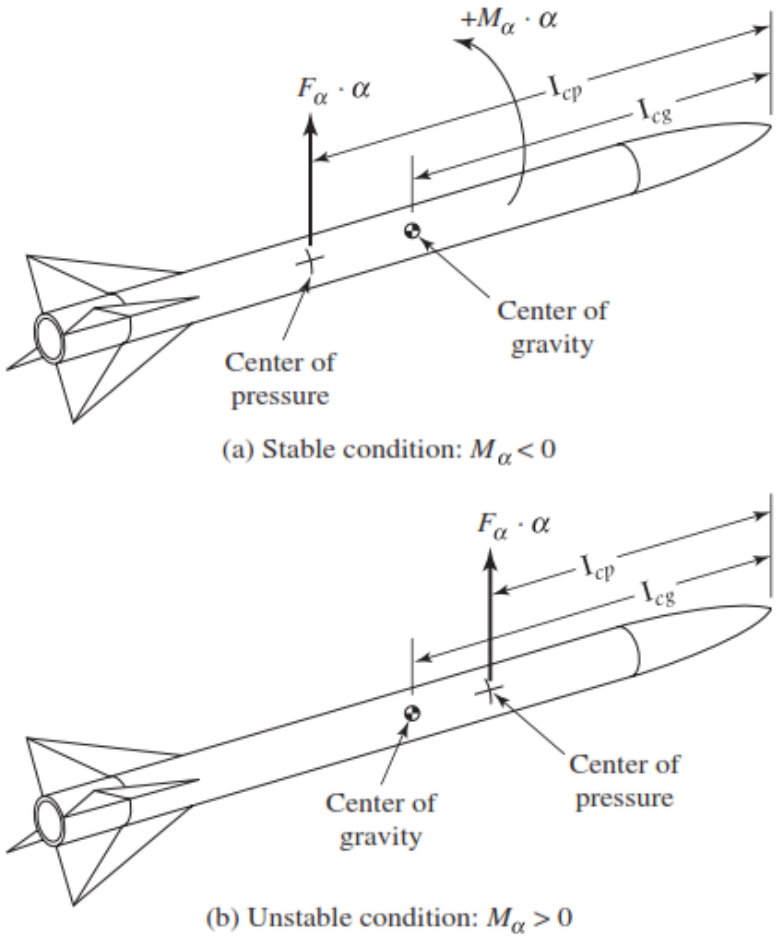


Figure 2.5: Illustration of a: (a) statically stable and (b) statically unstable missile airframe [244]

If an aircraft is longitudinally stable, a small increase in angle of attack will cause

the pitching moment on the aircraft to change so that the angle of attack decreases. Similarly, a small decrease in angle of attack will cause the pitching moment to change so that the angle of attack increases. M_α is defined as:

$$M_\alpha = \frac{1}{2}\rho V^2 \frac{Sd}{I_{yy}} C_{m_\alpha} \tag{2.40}$$

The same analysis can be done in the lateral axis (i.e. side-slip stability), in order to create a stable airframe [89].

2.7 Guidance Laws

Guidance is the strategy employed in order to generate an intercept with its intended target, expressed as a mathematical function. The mathematical expression calculates and acceleration command, $\underline{\mathbf{a_c}}$, which is the theoretical acceleration required to steer the missile to its target via the guidance strategy employed. This section discusses two classical approaches outlined below.

2.7.1 Line of Sight Guidance

Guidance can be defined as the strategy to steer the missile to its intercept. In the case for line of sight (LoS), the guidance algorithm aims to steer the airframe onto the line of sight, towards the target from its initial start/release point, as illustrated in Figure 2.6. It is important to note that if the guidance commands to follow the line of sight are generated externally and transmitted to the missile, then it is known as command to line of sight (CLOS) guidance [146].

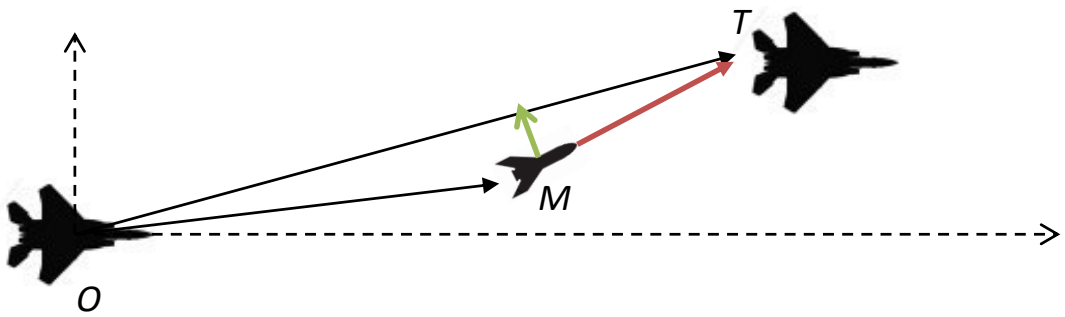


Figure 2.6: Illustration of LoS guidance

O is the initial release point; M is the guided object, which aims to travel along the line of sight vector, **OT**, towards the target, T. The LoS guidance aims to minimise the angle between **OT** and **OM**, as represented by the green line in Figure 2.6. LoS

is also referred to as three point guidance as there are three points of interest; 1) The launcher/tracker, O. 2) The guided object/missile, M and 3) The target, T.

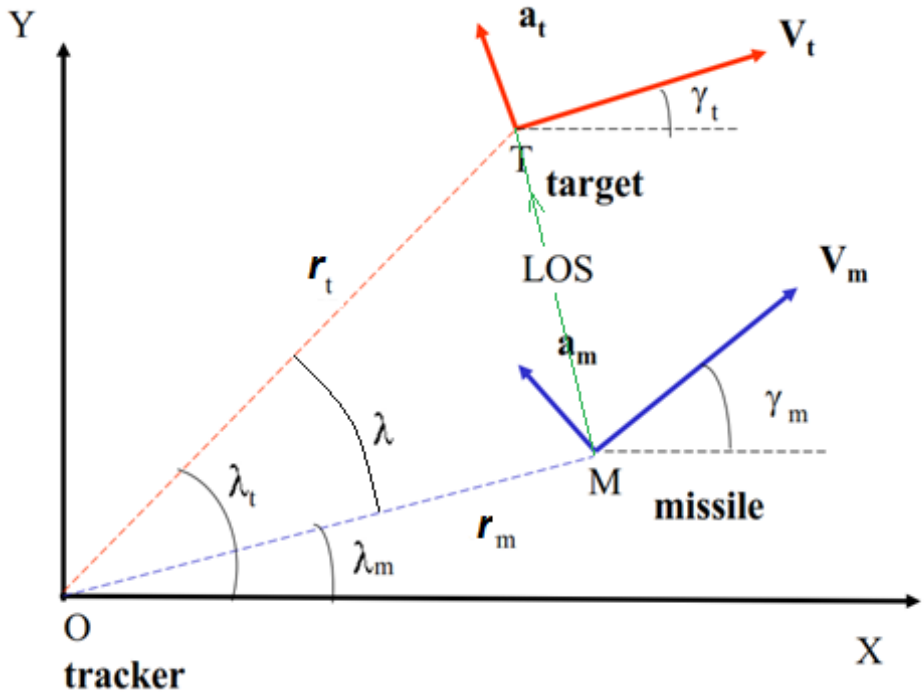


Figure 2.7: Detailed schematic of mathematical setup for line of sight guidance (adapted from [154])

In figure 2.7, λ_t is the angle from the line of sight from the target to the tracker, and an arbitrary horizontal reference line, commonly taken from the tracker's position. Similarly, λ_m is the angle from the line of sight from the missile to the tracker, and the same horizontal reference line. The line of sight between the target and the missile is shown by the green line in the figure above. The aim of the guidance law is to calculate the accelerations required in order to keep the distance from the LoS equal to zero or as close to zero as possible, throughout the time of flight of the missile. In three dimensions, the distance to line of sight line is known as the cross-range, and the aim of the guidance law is to minimise the cross-range error. The cross-range can be mathematically defined as: $(\underline{\mathbf{r_T}} \times (\underline{\mathbf{r_M}} \times \underline{\mathbf{r_T}}))$ [241].

The required guidance acceleration for LoS in three dimensions can be mathematically expressed as below [241]:

$$\underline{\mathbf{a_c}} = \frac{K\nu}{\left|\underline{\mathbf{r_T}}\right|^2}(\underline{\mathbf{r_T}} \times (\underline{\mathbf{r_M}} \times \underline{\mathbf{r_T}})) \tag{2.41}$$

2.7.2 Proportional Guidance

Proportional navigation (PN) guidance is another form of guidance law. PN guidance aims to intercept by keeping the rate of the angle between r_T and r_M, $\dot{\lambda}$, constant.

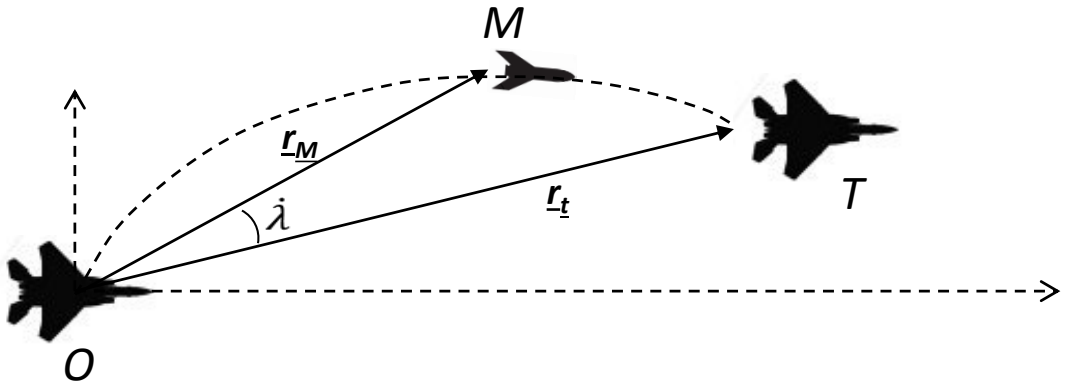


Figure 2.8: Illustration of PN guidance

The acceleration command is proportional to the rate of change of the line of sight angle and can be expressed as [203]:

$$\underline{\mathbf{a_c}} = N' V_c \dot{\lambda} \tag{2.42}$$

N' is the effective navigation constant, V_c is the closing velocity between the interceptor and the target. PN is usually the most commonly used guidance law due to its simplicity, effectiveness and ease of implementation [296]. There are many different forms of proportional navigation examples of these are: pure proportional navigation (PPN), this is the very basic form of PN guidance and can be mathematically expressed in 3D vector form as:

$$\underline{\mathbf{a_c}} = N'(\underline{\boldsymbol{\omega}} \times \underline{\boldsymbol{\nu_i}}) \tag{2.43}$$

where $\underline{\boldsymbol{\omega}}$ is the angular velocity of the *LoS* vector, r, between the target and the interceptor, and $\vec{\nu_i}$, is the velocity of the interceptor/missile.

Another form of PN guidance is known as true proportional navigation (TPN) and can be expressed as:

$$\underline{\mathbf{a_c}} = -\frac{N' V_c}{r} \nu_{c_\perp} \tag{2.44}$$

where r is the *LoS* vector as previously mentioned and $\nu_{c_\perp}$ is the closing velocity vector, normal to the r vector.

Augmented PN guidance accounts for targets with a constant accelerations and the effect of gravity and can be represented in vector form as:

$$\underline{\mathbf{a_c}} = N' V_c \dot{\lambda} + \frac{N'}{2} a \tag{2.45}$$

where a is the constant acceleration term given as $a = a_t - g_\perp$, a_t being the target acceleration and $g_\perp$ being the component of the acceleration due to gravity, perpendicular to the line of sight [212, 296].

2.8 Controls

Aerodynamic vehicles can have many types of control surfaces to steer the vehicle to its desired destination. Most conventional aircraft tend to have **canards** which are small surfaces located at the front of the vehicle near to the nose of the body. **Wings** are generally large surface located around the middle of the body but can vary depending on the desired performance parameter and **fins/tailfins** are small surfaces located at the rear of the body. There are other types of control surfaces and configurations available but the three described here are the most fundamental and most commonly utilised. The choice of control surfaces and there specific position(s) are dependent on the design considerations of the aerodynamic vehicle and its performance requirements [27].

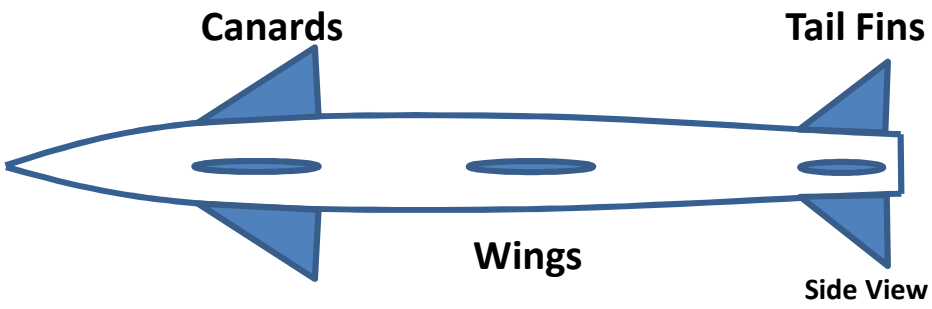


Figure 2.9: Conventional aerodynamic control surfaces

Depending on the design considerations there are two main types of turn control schemes used to steer an airframe. These are **skid-to-turn** and **bank-to-turn** [88]. Bank-to-turn implies that the airframe is banked at an angle relative to the wind, causing a component of the lifting force to act horizontally thereby causing the airframe to turn in the horizontal direction. Skid-to-turn refers to using the rudder or tail fins to apply a control surface deflection, in turn causing the airframe to "skid" in the intended direction of travel. In general, most aeroplanes will implement a bank-to-turn approach but typically small aerodynamic vehicles (such as missiles) tend to employ skid-to-turn. Skid-to-turn vehicles benefit from quicker turns, which is useful when trying to correct for small errors, but for large turning angles skid-to-turn tends to be less efficient and cause a greater loss of aircraft kinetic energy compared to bank-to-turn aircraft [27, 89].

There are two possible fin configurations that can implemented on missiles. The

configurations are known as "×" and "+" and the four control fins adjust accordingly to the specific manoeuvre desired (see Figure 2.10). There are both advantages and disadvantages to both configurations. The "+" configuration is a more simple control system as the fins are horizontal and vertical and they act in the same axes as the airframe. In terms of performance this fin arrangement tends to generate less drag force at trim conditions but often is less statically stable and often suffers from a statically unstable rolling moment derivative as compared to the "×" configuration. The "×" configuration is a more complex system to integrate into an aircraft, but has better roll stability, normally generates lower drag forces at most flight conditions and tends to be more suitable for launch platform integration.

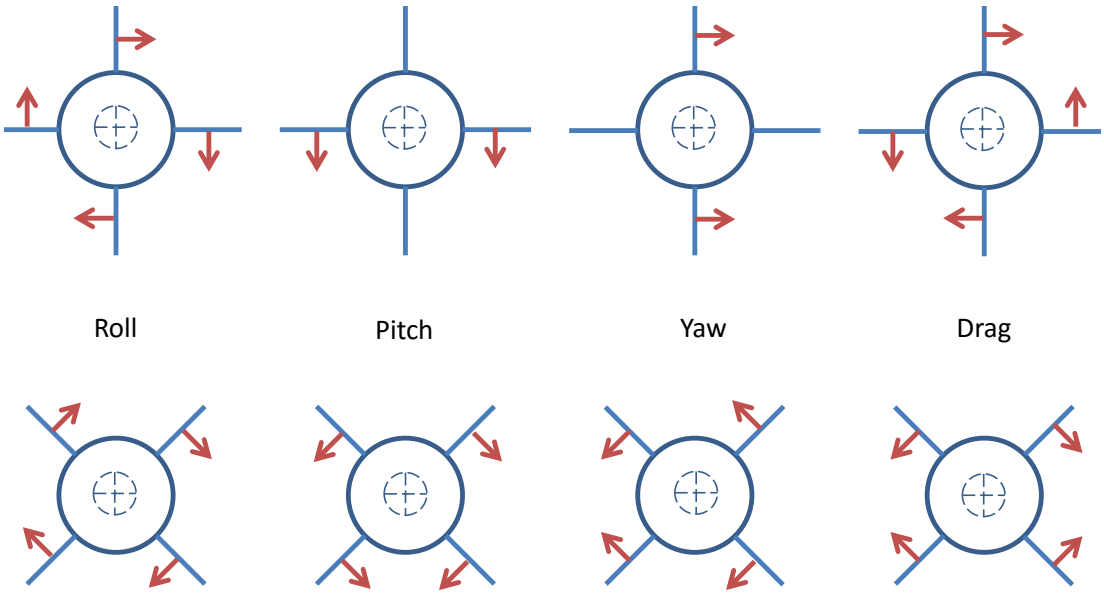


Figure 2.10: "+" (Top) and "×" (Bottom) control configurations. The red arrows indicate the direction in which the control surfaces deflect in order to generate the correct manoeuvre

2.9 Implementation

The 6DoF simulation model was created to describe the position, orientation, velocity and accelerations of a generic missile model. This will allow to generate trajectory data, in order to understand the camera motion the $dFoV$ imaging system will be subjected to. Using future air-to-surface guided weapon (light) (FASGW(L)) as the category of choice [261, 264], research and data was collected on existing weapons within this category and a generic model was created.

FASGW(L) was selected because it provides the most difficult scenario an imaging system must face. These weapon systems tend to be lightweight, short-to-mid-range, high velocity systems capable of precision strike. Its popularity in military applications

has grown considerably recently, alongside the Laser beam riding guidance system which will be detailed in the next chapter.

A form of guidance system (outlined in Chapter 3) is required to direct the missile towards its target (i.e laser beam riding). In the cruise/terminal flight phases an imaging seeker would be ideal for target detection and engagement. This is were the $dFoV$ imaging system would be beneficial.

The FASGW(L) performance requirements provide a difficult environment for an imaging system as the short-range, high velocity characteristic would require fast and precise control inputs to steer the missile which, in turn, produces undesired camera motions. This will allow to determine the motions an imaging system will undergo in a worst case scenario.

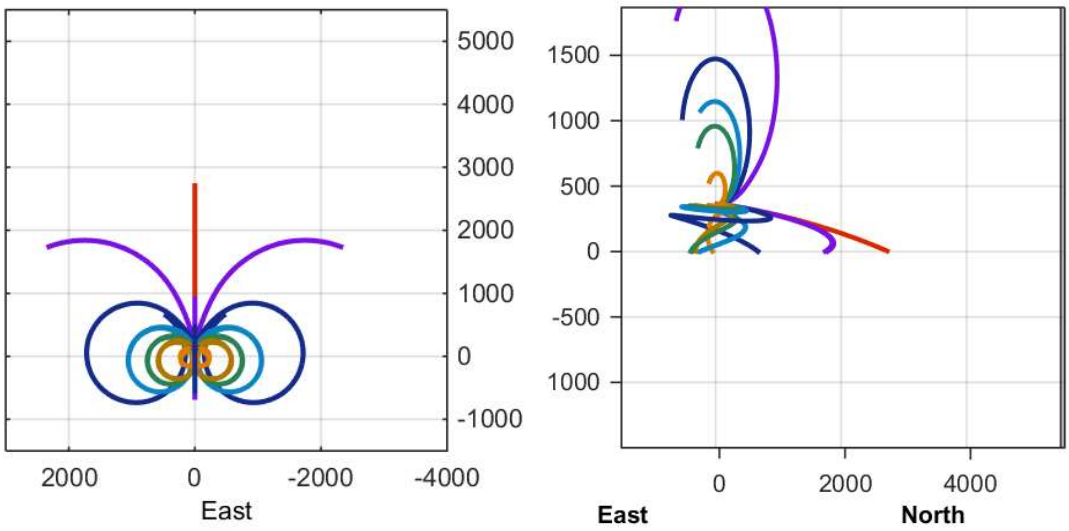
The characteristics of the designed missile model are presented in Table 2.1 and was based on current proposed FASGW(L) missiles proposed such as the Lightweight Multi-role Missile [261].

Table 2.1: Missile Model Parameters

Missile Parameters	**Values**
Body Length (m)	1.3
Body Diameter (m)	0.076
Canard Length (m)	0.125
Canard Width (m)	0.0435
Fin Length (m)	0.135
Fin Width (m)	0.0543
Initial Total Mass (kg)	13
Total Fuel Mass (kg)	3.3
Boost Fuel Mass (kg)	2.5
Cruise Fuel Mass (kg)	0.8
Top Speed (Mach #)	1.5
Typical Range (m)	8000

A PID (proportional-integral-derivative) controller [182, 183] was also implemented in order to stabilise the control deflections applied to the airframe. Illustrations are also provided describing the individual subplots in the model (see Figure 2.11).

Following on from this, a test on the "targetable region" was carried out which shows a 3D envelope of the area in which the missile is able to hit the target within, by applying a certain amount of control deflection in the pitch and yaw axes. The figures below shows the outcome.



(a) Top-view

(b) Side-view

Figure 2.12: Illustration of the top-view (a) and side-view (b) of the missile's manoeuvrability envelope

Figure 2.11 below, shows the results of the moving target with constant velocity, with a trajectory plot in the XZ plane. The figure shows the graphical user interface created for analysing the performance, accuracy and stability of the missile, from its release to when it engages with the target. The model is able to use the different forms of guidance laws described earlier (i.e. LoS, PN and LBR (discussed in the next chapter)). The position and velocity of the target, and the missile, can be modified for various scenarios. Further details of the individual plots generated, have been illustrated within the figure.

Figure 2.12 illustrates the manoeuvrability and stability of the designed missile model. Figure 2.12b shows a side on view showing the possible pitch trajectories with a constant deflection ranging from 0° to 10. Figure 2.12a show a top view, illustrating the possible yaw trajectories with a constant surface deflection ranging from -10 to 10. The figure illustrates that even through large control deflections, the missile is able to follow its flight path without becoming unstable and uncontrollable, whilst also maintaining its manoeuvrability.

Below, Figure 2.13 shows the trajectory plot from the XY plane showing the target moving to the right and the missile pursuing the target. The blue line illustrates the flight path travelled to reach the target. The moving target was modelled using the kinematic equations with constant velocity (no acceleration).

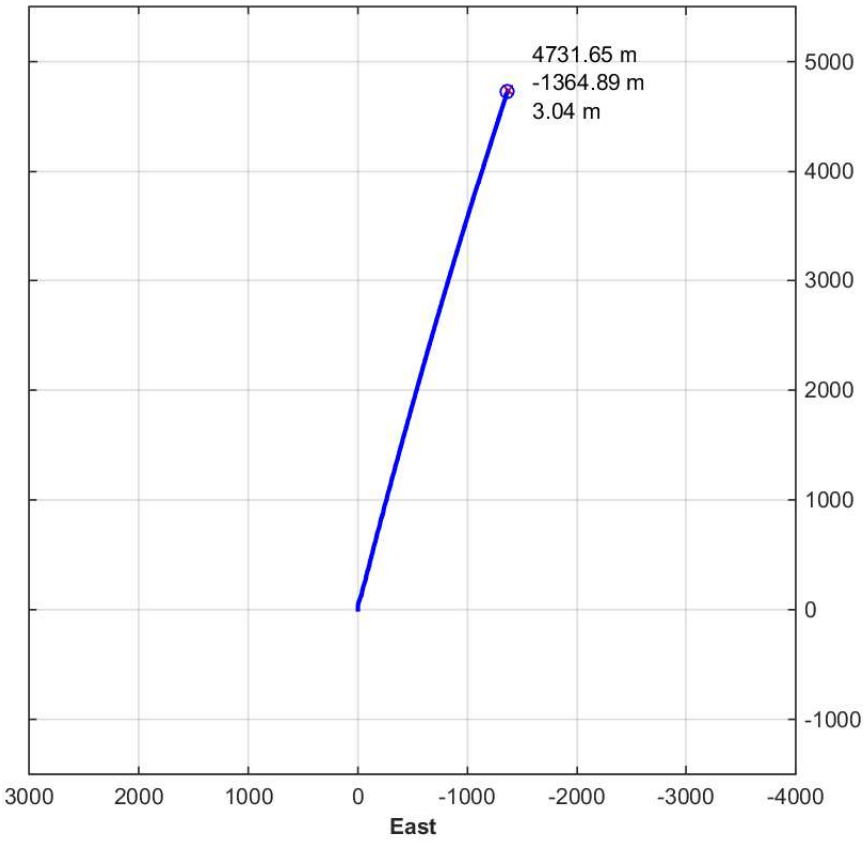


Figure 2.13: Moving target and missile, with trajectory path in XY plane

The accuracy is within $4m$ which is within the region of the operationally relevant miss distance of our generic missile model ($= 10m$). This could be improved further by using a different guidance law. The LoS guidance is always following the missiles current position and not predicting the future position, which would generate a higher accuracy of an intercept.

Another factor to consider is the stability of the generic missile, as if the missile is unstable, it becomes very difficult to use the guidance laws to track and steer towards the target. Therefore a simple longitudinal static stability analysis was carried out as described in Section 2.6, which in essence, states that $\frac{dM}{d\alpha} < 0$. Figure 2.14 below shows a plot of the total moment co-efficient against the angle of attack.

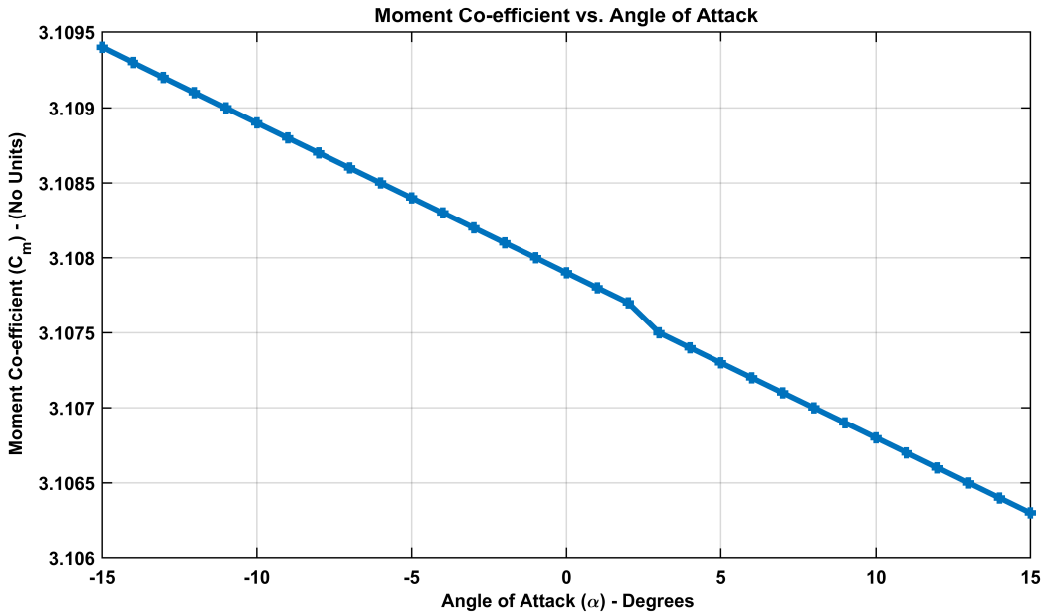


Figure 2.14: Longitudinal stability analysis - plot showing moment co-efficient (c_m) vs. angle of attack (α)

The gradient of Figure 2.14 represents $\frac{dM}{d\alpha}$, more accurately $\frac{dC_M}{d\alpha}$. The negative gradient shows that the missile model is statically stable in the longitudinal direction. The gradient was calculated to be -1.04×10^{-4}, which is a small negative value, therefore the missile is stable, controllable and manoeuvrable.

2.10 Chapter Summary

A FASGW(L) specification 6DoF missile model was designed and developed, using current theory available. The performance, stability and manoeuvrability of the model was assessed to understand its capabilities and limitations. A PID controller was also implemented alongside the LoS and PN guidance laws. Chapter 3 will discuss the available guidance systems and seekers available to implement in conjunction with the guidance laws, detailed in this chapter. It is worthy to note that the theory used to implement the 6DoF model is widely available, but the implementation of a FASGW(L) missile model, on the other hand, is something that has not been conducted outside of military research, testing and validation.

The missile model integrated with the guidance system developed in Chapter 3, will provide the necessary trajectory data required to develop an understanding of the camera motions (i.e. pan, tilt, zoom etc.). The *dFoV* imaging system will typically undergoes during operational use, this will help evaluate the image separation technique described in Chapter 4.

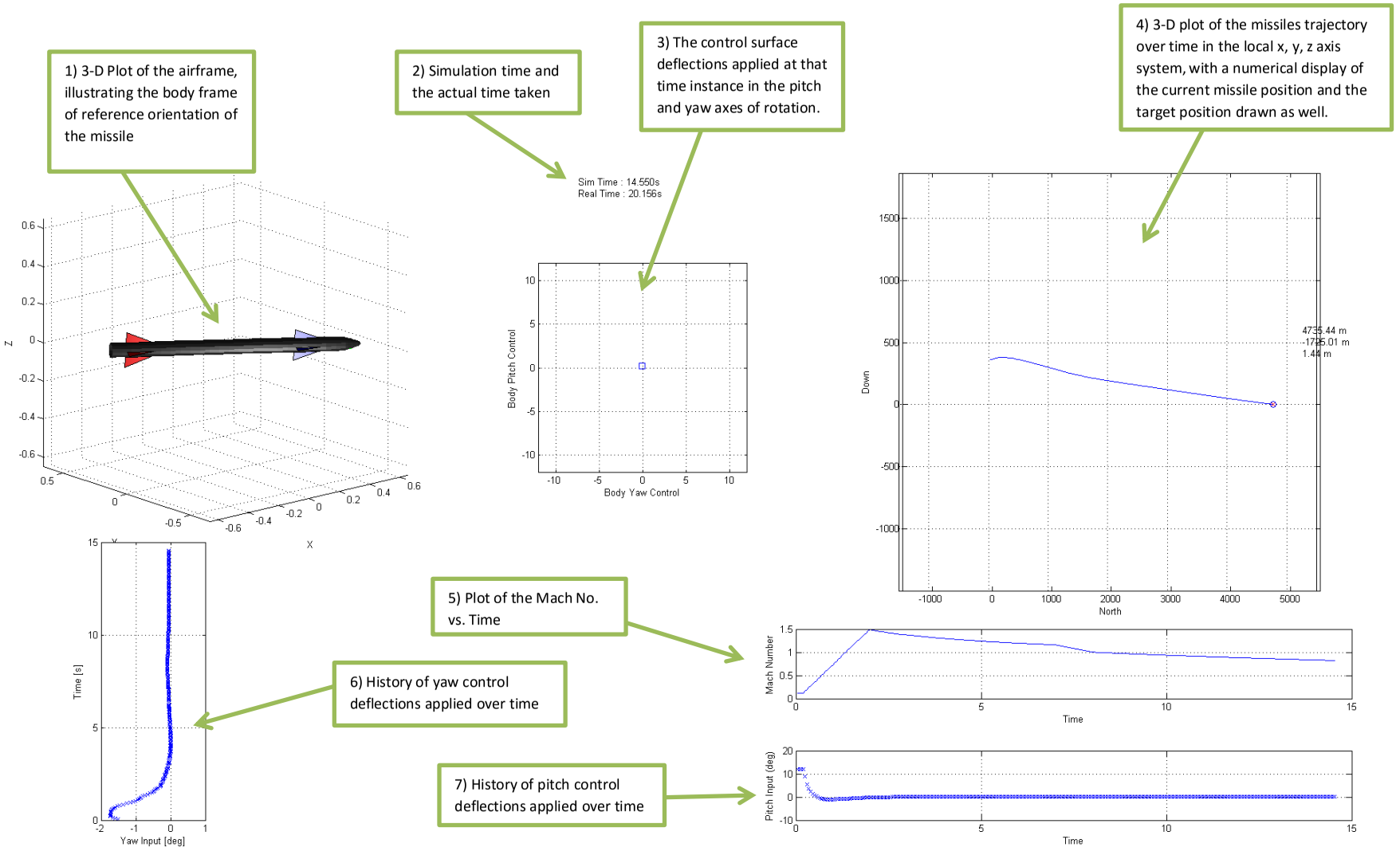


Figure 2.11: MATLAB simulation model graphical user interface (GUI)

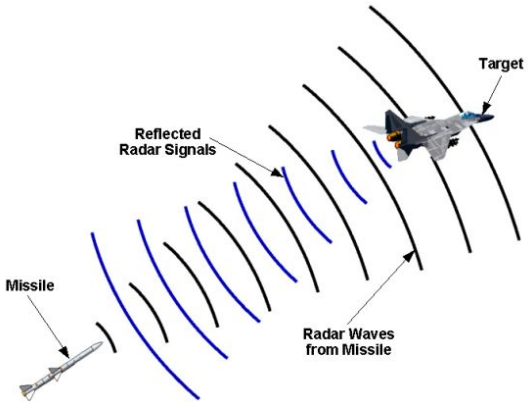
Chapter 3

Guidance Systems and Seekers

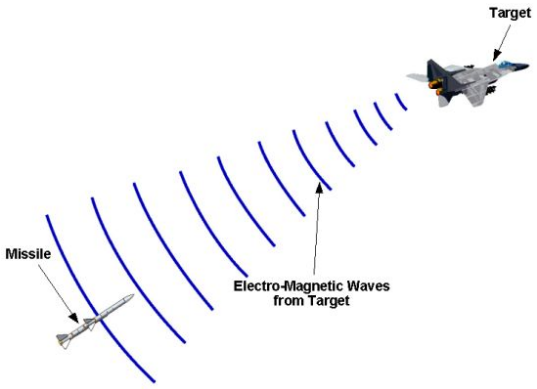
Seekers, are sensors that generate the measurements for target detection and homing, by sensing, radio frequency (RF), infrared (IR) or visible energy that the target emits or reflects and is then used as part of the overall guidance system to generate an intercept. A seeker can be considered as the eye of a missile, sometimes called the "homing eye". Examples range from synthetic aperture radar (SAR), heat seekers, millimetre wave radars, electro-optical sensors and even laser beam riding systems.

3.1 Types of Guidance Systems

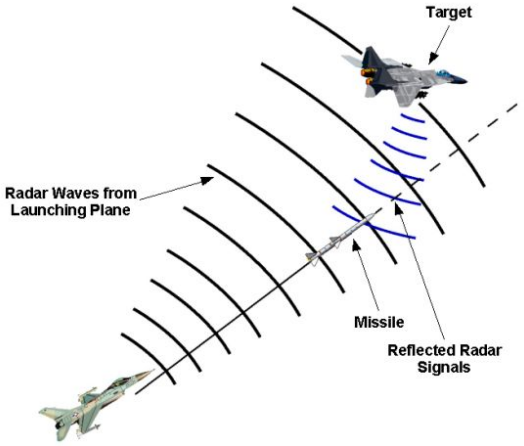
There are three main types of guidance systems: passive, active and semi-active seekers, (they can also be categorised into imaging and non-imaging seekers). Passive seekers typically detect and track objects by processing reflections from non-cooperative sources of illumination, such as communication signal or heat signatures generated by aircraft, such as the hot exhaust gases (Figure 3.1b). Active seekers typically are on-board the airframe and provide their own guidance signal. Typically, the systems aim to keep the airframe pointed directly at the target, and attempts to minimise the error between the angle from the guidance signal and the angle of the airframe from its centreline (Figure 3.1a) [69, 88]. Active seekers are widely used in anti-ship missiles and in air-to-air missiles and is often termed fire-and-forget, examples of such missile systems include AIM-120 AMRAAM and the AGM-65 Maverick and its many variants. Passive systems tend to produce accurate angle information but inaccurate range estimates, whereas active seeker systems tend to produce accurate range information but have coarse angular resolution. Semi-active seekers combine a passive receiver on-board the airframe with a separate targeting device (transmitter), that illuminates the target (Figure 3.1c). The illumination can be visible or non-visible forms of radiation, in the form of a laser or radar, and can be encoded to prevent unauthorised interception of the signal [125]. The figure below (Figure 3.1), illustrates the different methods each seeker type uses.



(a) Active seeker



(b) Passive seeker



(c) Semi-active seeker

Figure 3.1: Illustration of the different seeker types, (a) active, (b) passive and (c) semi-active, obtained from [232]

Traditionally, seekers have been simple correlation trackers and a set of gimbals that are used to measure the angle rates of the target. For example, an infrared seeker detects and tracks the local hotspot on the target such as the exhaust plumes of an aircraft. A typical seekers hardware consists of a two or three gimbal mounted gyroscope as well as transmitter and/or receiver. Most seekers tend to contain a two gimbal axes gyroscope (yaw and pitch) and rely on the missile roll autopilot for roll stabilisation [244].

The main generic roles of a seeker are [244]:

1. Provide measurements of a targets motion, to allow the guidance law to function.
2. Acquire and track the target, using some form of energy receiving device e.g. radar, laser, infrared.
3. Measure the LoS (line-of-sight) angle rate ($\frac{d\lambda}{dt}$).
4. Stabilise the seeker against the missiles pitching rate ($\frac{d\theta}{dt}$) and yawing rate ($\frac{d\phi}{dt}$).

3.1.1 Spinning Reticle Seeker

A spinning reticle seeker is an optically modulated seeker. It performs two important functions, providing the guidance system with directional target information and suppressing background IR radiation. A spinning reticle seeker contains a infrared detector and a spinning pattern disk containing a particular pattern of opaque and transparent areas on its surface [147]. The AIM-9 Sidewinder missile used an IR rotating reticle seeker [209, 286]. An example of such a seeker is the "Rising Sun" reticle seeker [46].

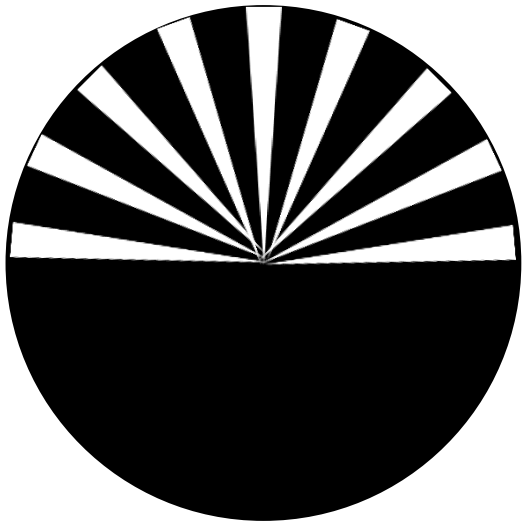


Figure 3.2: "Rising Sun" spinning reticle seeker

If the target hotspot is at the centre of the disk, and the rotation of the disk is constant, this means the missile is pointed at the target. When viewing the signal of the thermal detector generated from this seeker and in this scenario the output would be a constant. If the target is moved off centre, the signal output is modulated by the bars and zero when the opaque semi-circle is covering the target hotspot (See Figure 3.3). As the bars are slices of the circle itself, the frequency of the modulation does not change as the target hotspot move towards or away from the centre. Instead the phase of the pulse-null signal waveform provides an indication of the direction of the target hotspot. The reticle pattern of the "Rising Sun" seeker, has one disadvantage in that it is unable to detect the radial position of the target hotspot.

Figure 3.3: Output signal from thermal detector of "Rising Sun" spinning reticle seeker

There are many other variants of the spinning reticle seeker, with more advanced reticle patterns [72, 86] which can provide more positional information of the target hotspot such as the once shown in Figure 3.4, but this is out of the scope of the thesis.

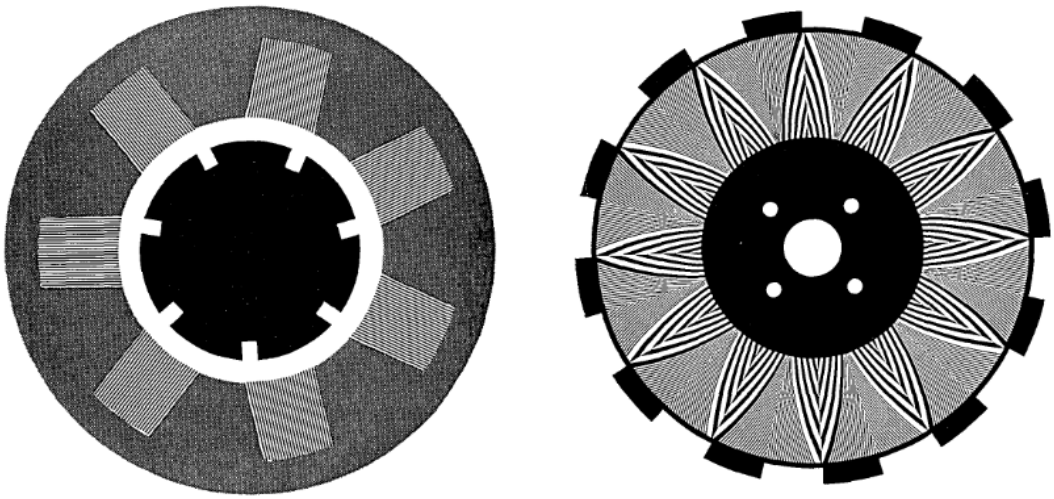


Figure 3.4: Examples of more advanced spinning reticle seekers (obtained from [86])

3.1.2 Quadrant Seeker

Another variety of seekers for guidance systems is the quadrant seeker. An example of a weapon system that uses this seeker is the Paveway II GBU [268, 269]. This is a simple model which consists of a detector (for example, thermal or laser detector), split into four sections, with a central ring, which when the hotspot is in, the missile is pointing towards the target hotspot. The quadrant seeker is shown in Figure 3.5.

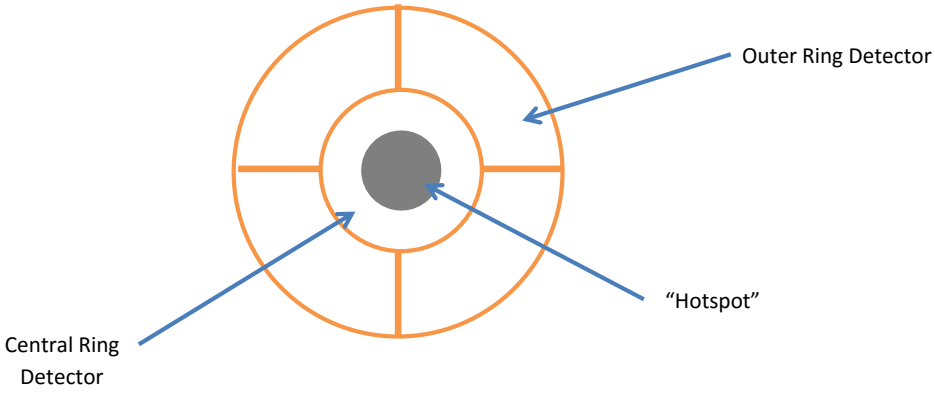


Figure 3.5: Quadrant seeker

In a quadrant seeker, all four quadrants of the seeker are equally illuminated. If the hotspot lies off -centre, it will illuminate each quadrant differently. Each quadrant generates an electrical current that is proportional to its illumination. Voltages pro-

portional to these currents are then amplified and fed into a mixer, which compares the signals and generates the appropriate guidance commands to bring the hotspot back to the centre [146].

3.2 Beam Riding Guidance Systems

An increasingly popular guidance system is the beam riding system. The beam riding systems operates by focusing some form of focused electromagnetic (EM) radiation, at an intended target. The two most common utilised forms of beam riding systems are radar and laser beam riding systems. Beam riding systems operate by pointing a signal beam towards its intended target, and after the launch phase, the missile "rides" along the beam, until it has intercepted the target. There are certain advantages of using a beam riding system, for example, the missile rides the beam, and does not rely on the EM energy reflected or generated from the target itself, and therefore makes it much more difficult for the target to deploy a countermeasure or attempt manipulate the signal.

3.2.1 Radar Beam Riding

Radio detection and ranging, or commonly known as radar, is a system that was developed during World War II. Radar was developed to detect objects such as ships, aircraft and missiles by using radio waves. Radar beam riding, is the semi-active guidance system which uses radio waves as a beam to direct the missile to its desired target. The radar beam's EM properties are modulated in specific ways to allow the missile's guidance system to interpret its location with respect to the beam centre.

A conventional radar guidance system typically operates using an antenna to transmit a series of EM radio waves, and an antenna to detect the EM waves that have been reflected off the intended target. Typically, the same antenna can be used to both transmit and receive signals. The time delay between transmission and receiving of the reflected signal can also provide range information.

For a radar beam riding guidance system, the transmission antenna is directed to the intended target as expected but the receiving antenna is mounted on the missile itself and does not rely on the reflected waves to return to the transmitting antenna. It also prevents the target from deploying countermeasures to jam, deflect or absorb the transmitted signal, to prevent the reflected signal from reaching the receiving antenna.

The missile computer system compares the signal strength from different locations on the missile body and steers the missile to maximise the strength of the beam, which is at the centre of the beam [30]. Radar beam riding systems also employ a conical scanning technique, in which the radar beam is rotated in a circular manner around the boresight axis (beam centre) aimed at the target, to allow the missile to more

accurately locate the beam centre.

3.2.2 The Bumblebee Project

Following the details outlined in [118], in 1925 an early radar missile guidance system was proposed, equipped with photo cells on the tail fins. The rocket would be launched into a searchlight beam and would follow the beam to the illuminated target. The system was never produced but was the basis of the beam riding guidance system for the early Bumblebee missile system. The Bumblebee guidance concepts were based on pulsed radar technology [30, 112, 118, 202, 287].

For long range targets, some form of terminal homing was required to intercept the target, due to the limited range of radar systems. Initially, the target radar beam functioned as the guidance beam, to steer the missile using line-of-sight guidance, but this was inefficient for crossing targets that fly perpendicular to the line of sight and eventually a more advanced system which used proportional navigation was implemented, but the missile was still required to stay within the beam, causing the missile to fly an inefficient path [112].

The most effective use of the missile, for long range targets, was to launch it to a high altitude cruise and then guide it to an intercept point ahead of the target, to eliminate any form of 'lead-lag' (i.e. where the missile is continuously 'chasing' the target) [13]. This scheme required two separate radar beams for target tracking and missile guidance, as illustrated in Figure 3.6.

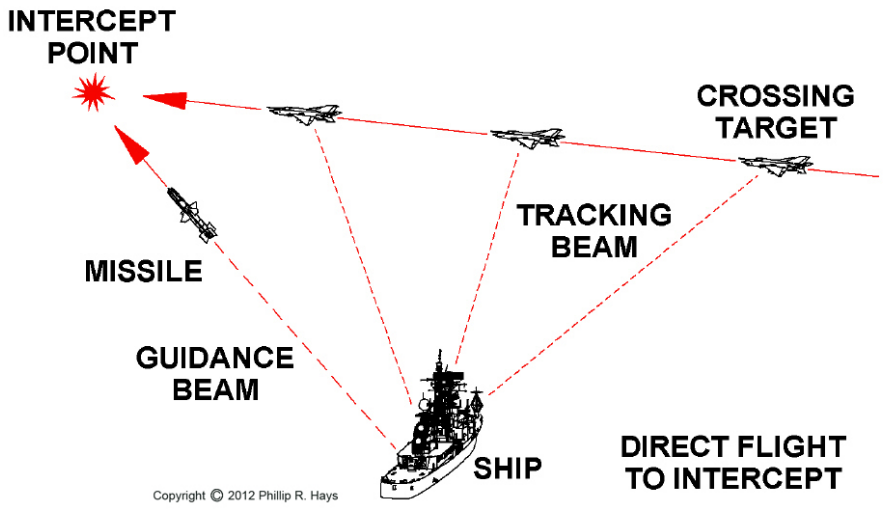


Figure 3.6: Direct flight to intercept [118]

Using this scheme, with two radars, allowed a fire control computer to obtain a target's position and missile position information from the tracking radar and the missile guidance radar respectively. This, in turn, allowed the fire control computer to calculate an intercept point, and then steer the guidance beam to deliver the missile to

the target.

As an example, the Talos missile is a long range ship-borne surface-to-air missile. Its guidance beam would rotate to sweep around a conical scan with respect to the direction of the ship to the missile. This created a ring pattern of different intensity of energy, as illustrated in Figure 3.7.

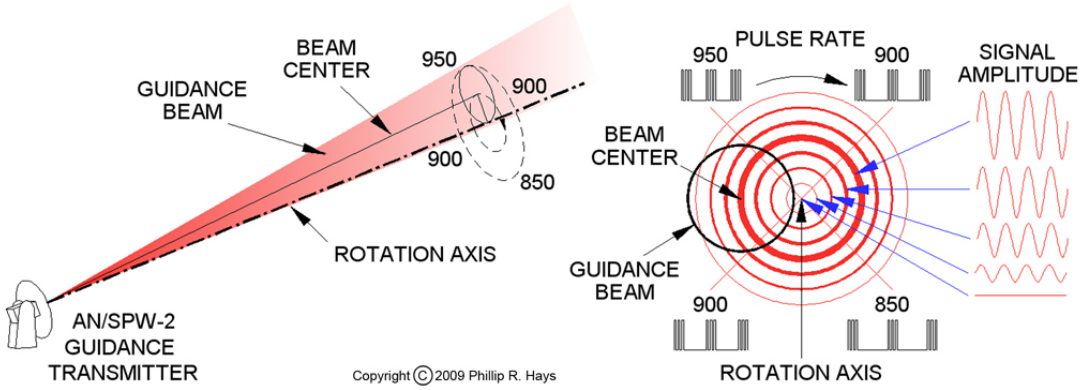


Figure 3.7: Conical scanning radar beam guidance [118]

Following the description in [112, 118], when the missile is in the centre of the rotation axis, of the guidance beam, it is receiving a signal with constant amplitude as shown in Figure 3.7 (right), but when it deviates away from the centre the signal amplitude starts varying in a sinusoidal form as illustrated in 3.7. The beam is rotated 30 times a second producing a 30Hz sinusoidal frequency modulated (FM) signal. If the missile is in the centre of the rotation axis, constant signal strength is received by the receiver, and no manoeuvring is required. If the missile was off-centre from the guidance beam it would create a sinusoidal signal as the beam rotates around the rotation axis at 30Hz. The timing of this AM signal with reference to an FM reference signal determined the position of the missile within the guidance beam, and the rate of the pulse groups determined the direction the missile needed to turn, in order to steer the missile back to the rotation axis centre. The pulse group was different for each guidance transmitter to keep the missile locked on to one guidance transmitter only, and prevent any interference between multiple transmitters working at the same time.

The main disadvantage of radar beam guidance was that the accuracy is dependent on the width of the guidance beam, which increases with distance from the radar, hence the accuracy of the target degrades. Radar signal are also very difficult to encode, to provide the positional information of the missile with respect to the beam centre as described above. More modern beam riding systems tend to use laser beams instead of radar, as they are simple to implement, reliable, compact, insensitive to distance and generally difficult to detect and jam [146]. As will be discussed in the following section.

3.2.3 Laser Beam Riding Guidance System

The name laser, originated from the acronym for light amplification by stimulated emission of radiation. Laser beam riding (LBR) systems operate in the same manner as radar guidance systems, except that the signal source is different. The laser beam, provides certain advantages such as a narrower beam width due to the shorter wavelength, a stronger signal with better range due to its coherence properties. Furthermore, The laser can be optically encoded without much difficulty as opposed to radar systems. These advantages are mainly due to the properties of lasers, which are outlined below.

Laser Light Properties

The process of stimulated emissions creates laser light with specific properties making it different from normal light. Laser light has the following properties [120, 285]:

1. The light released from the laser is monochromatic, meaning it contains one specific wavelength of light (colour) or a very narrow range of wavelengths and is dependent on the lasing medium and design of the laser cavity.

2. The light released is coherent. Each photon moves in phase with others. This means that all of the photons have a wave front that is in unison (see Figure 3.8).

3. In reality it is not perfectly coherent, and can easily be affected by the outside environment.

4. The light released is very directional, hence laser light has a very tight beam width (beam divergence) and is very concentrated, but for long range applications the beam divergence should be considered. Although semiconductor lasers naturally have a high divergence, external optics can focus their output into a pencil-like beam.

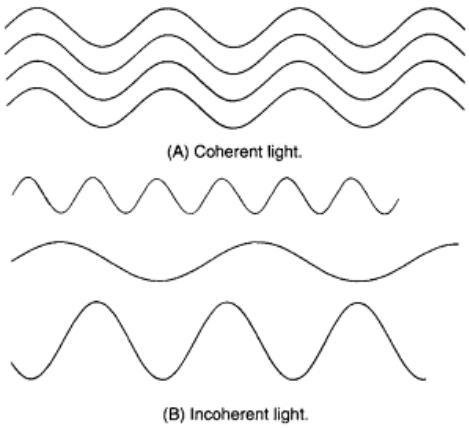


Figure 3.8: Illustration of (A) coherent and (B) incoherent light [120]

There are many different types of lasers such as solid-state lasers, gas lasers, excimer lasers, dye lasers and semiconductor lasers. For military purposes, it seems feasible that semiconductor lasers (also known as diode lasers) are the most practical form of lasers to use due to its small form factor, portability and relatively low power [120, 146, 285].

3.2.4 Types of Laser Guidance Systems

Laser guidance systems can range from active, semi-active, passive similarly radar guidance system as described earlier in Section 3.1, but more recently laser beam riding (LBR) systems have become popular for future military applications such as the Lightweight Multi-role Missile (LMM) by Thales [261], developed for high velocity (Mach ≈ 1.5) with an operational range of 8km, weighing only 13kg, designed for air-to-air (A2A) and surface-to-air (S2A) engagements, with a variety of launch platforms, for example Lynx Wildcat helicopter, man portable and tanks.

Another example is the Paveway GBU. A LBR system is a type of semi-active command-to-line-of-sight (SACLoS) guidance system, whereas the Paveway GBU is a form of semi-active laser (SAL) guided system. The main difference is that for a SAL guided weapon the seeker is located on the nose of the missile, and the energy reflected off the target is tracked in order to generate an intercept with the target. Whereas, in a SACLoS system the seeker is located at the rear of the missile, typically on the tail fin. The rear seeker keeps the missile on the beam, and follows it to the desired intercept. There are advantages and disadvantages to both systems for example SACLoS systems are less susceptible to counter-measures/jamming, SAL systems tend to be more accurate at longer ranges, although this is much less now, due to technological advances. The SAL guided Paveway is not a beam riding system, the seeker tracks the position of the target, using the reflected energy given off the target by the laser designator, and then uses proportional navigation to direct the weapon to the target. The difficulty lies in how the missile is performing this task. There are many possible options such as: pulse modulation (PM), frequency modulation (FM), amplitude modulation (AM), phase shifting or through the use of a seeker.

3.2.5 Laser Beam Riding System

A typical laser beam setup for an ATGM (anti-tank guided missile) is illustrated below.

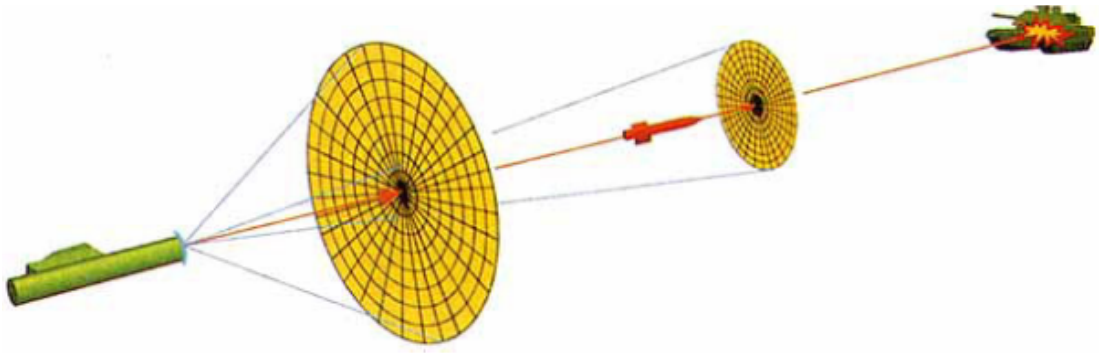


Figure 3.9: Schematic of laser beam riding system [154]

The laser beam is spatially encoded with a pattern to enable the missile to be able to sense its position relative to the LoS/beam centre. The spatial encoding of the beam allows the missile to generate accelerations to keep the missile on the beam centre internally. It should be noted that in a laser beam riding system the missile is blind as it does not see the target, as there is only a rear facing seeker, tracking the laser beam [154]. As the missile rides the beam, it will veer away from the beam centre, as this happens the spatially encoded laser pattern, will allow the on-board missile computer to deduce its position from the beam centre and then the LoS guidance will compute the required accelerations and hence control surface deflection required to steer the missile back to the beam centre.

3.2.6 Implementation

Following the FASGW(L) requirements, a laser beam riding guidance system was implemented into the current simulation model described in chapter 2. The laser guidance system will be broken down into several subsystems: the laser model which will model the source of the beam and the rear facing seeker onboard the missile. Initially, it was assumed that the laser beam will be an ideal pencil beam, hence it will be assumed that the beam divergence is zero. The backwards facing seeker is the most important sub-system as this will allow the missile guidance computer to calculate its position relative to the laser beam. For this model, a simple quadrant sensor will be used, as shown in Figure 3.5

As the missile veers away from the central ring, the "hotspot" i.e. the beam centre will come into contact with one, or more, of the outer ring detectors, depending on which 'quadrant(s)' the hotspot is detected in, the position of the missile relative to the beam centre is determined, from this the missile guidance computer will apply the corresponding controls to bring the hotspot back into the central ring and hence the

missile will once again 'ride' the beam, the hotspot is monitored throughout the flight of the missile.

The LBR system is implemented to be used for the mid-course guidance i.e. the cruise phase of the missile flight, it will steer the missile towards the general direction of the target. In the cruise flight phase, the $dFoV$ imaging system can be activated to detect and track the target, once the target has been acquired, and the missile will intercept the target following its positional information provided by the $dFoV$ imaging system.

In the next section, a review of the current available imaging seekers will be conducted, and the proposed $dFoV$ seeker will be outlined.

3.3 Imaging Seekers

Non-imaging seekers use radiation emitted from, or reflected off the target, to detect and track aircraft. Imaging seekers operate not only in the temporal domain but also in the *spatial* domain [65]. This can be achieved by using a photodetector array often referred to as a focal plane array [76]. Imaging seekers, similar to non-imaging seekers can operate across the electromagnetic spectrum at different operational wavebands (i.e. visible, infrared, TV).

An imaging seeker system often contain the following sub-components:

Optics i.e. Aperture, lenses, optical filters.

Photodetector e.g. 2D array of photodiodes [67].

Read-out circuits i.e. Integration circuit, gain-offset correction and analogue-to-digital (A2D) converter.

Storage e.g. Electronic memory.

3.3.1 Visible

Visible band image seekers/cameras operate in the $0.4\mu m$ to $0.7\mu m$ waveband, these systems are generally good, as they are in high demand, not only for military applications but also in commercial usage, making them relatively cost-effective to manufacture. Visible band cameras tend to suffer from poor performance in low-light conditions (e.g. at night), this can be mitigated using a "low light" system which amplifies the light before the image is formed, but such systems require extra signal processing and tends to introduce noise as well as making them sensitive to glare. Visible band cameras are often used in television guidance missile (TGM) systems, which uses a visible band camera onboard the missile. The signal is transmitted to the remote terminal operator (RTO), from which the missile is steered towards its target. An example of such a weapon system was the AGM-62 Walleye glide bomb used by the US armed forces

during the 1960s, more recently it has been superseeded by variant-A of the AGM-65 Maverick and now replaced by variant-D of the Maverick which operates using imaging infrared (IIR) system.

3.3.2 Infrared

Infrared (or heat-seeking) imaging seekers [147], often called imaging infrared (IIR) seekers or thermal imagers, use the infrared electromagnetic spectrum i.e. $0.7\mu m - 14\mu m$ where there are large amounts of ambient light due to thermal emissions (e.g. jet exhausts). Infrared imagers operate in three sub-bands: near-wave infrared ($0.7\mu m$ to $\sim 2\mu m$), mid-wave infrared ($3\mu m$ to $\sim 5\mu m$) and long-wave infrared ($8\mu m$ to $\sim 14\mu m$). Infrared imagers tend to be difficult to manufacture, as standard silicon photodetectors cannot be used, the focal plane array usually requires some form of cooling for the imager to operate optimally and there is less demand for them, making them expensive than visible band systems. An example of a weapon system which utilises infrared imaging systems is the AIM-9 Sidewinder, which is a short-range air-to-air (A2A) missile. It began service in 1956 with several variants released for different operational requirements but most of the major variants included a IR seeker such as the AIM-9X [148, 191].

3.4 Dual Field-of-View Imaging Seeker

In this section, a novel imaging seeker or camera is proposed. In operational use on-board aircraft systems, in particular missiles and unmanned aerial vehicles (UAVs). Imaging systems are required to provide situational awareness and target detection, identification and in certain cases, target tracking capabilities. These operational requirements often come as a trade-off. Situational awareness is required in order to surveil a large area and hence a wide field-of-view ($wFoV$) camera is required, but in order to detect, identify and track possible targets, a narrow field-of-view ($nFoV$) camera which provides high spatial details of a small area is necessary.

It is with these opposing operational requirements that an alternate solution is proposed. Current systems would have two (or more) separate imagers ($wFoV$ & $nFoV$) which operate in a coordinated manner, such that the $wFoV$ imager will cue the $nFoV$ camera to interrogate possible targets. Examples include the LANTIRN [174] and LITENING [110] targeting pods, designed by Lockheed Martin and Northrop Grumman respectively. For lightweight UAVs, this can be a limiting factor, due to payload and performance, hence multiple camera systems can be impractical.

This thesis aims to discuss an alternative solution, in which only a single sensor is utilised to capture both the $wFoV$ and $nFoV$ simultaneously, the images captured will be superimposed onto each other, resulting in a superimposed image, which will

be separated using post-processing using a geometric relationship between the two images as proposed in the following chapters of the thesis, in particular chapters 4 & 6. Thus allowing for both situational awareness or surveillance ($wFoV$), as well as target detection, identification and tracking ($nFoV$).

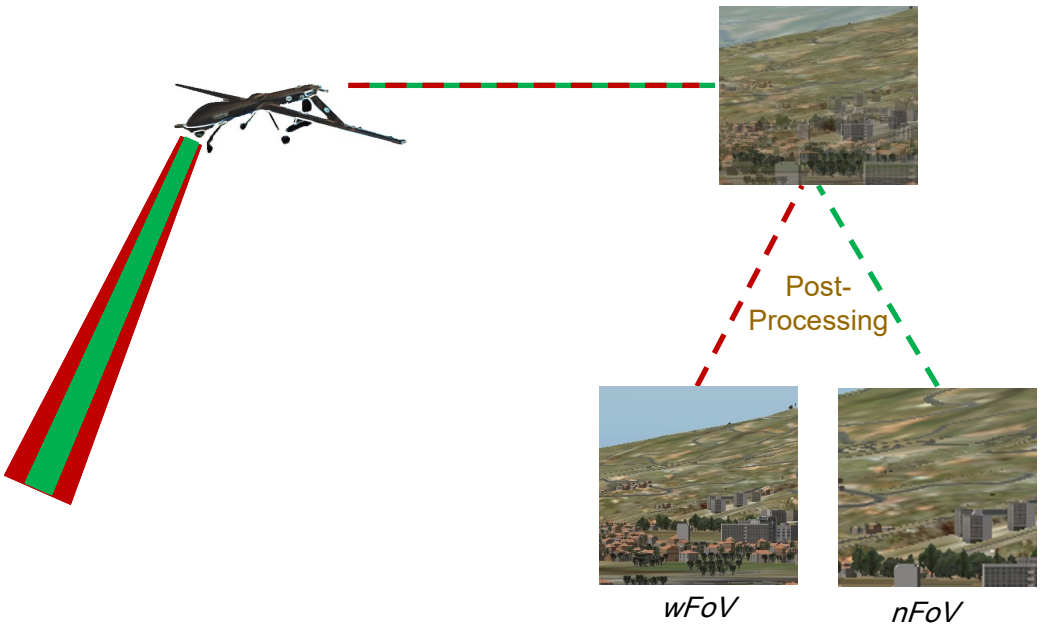


Figure 3.10: Illustration of the dual field-of-view imager scenario

3.5 Chapter Summary

In this chapter several state-of-the-art seekers and guidance systems required to detect and track targets have been reviewed. A novel approach to imaging using a dual field-of-view sensor has also been proposed and the scenario illustrated (Figure 3.10). The functioning of the seeker plays an important role in the overall missile system. The simulation model presented in Chapter 2, can only model the trajectory of an aerodynamic airframe. Incorporating a guidance system, allows to track a specific target, even if the target is moving. This is a much more realistic environment, and facilitates to generate realistic trajectories and behaviour, which will generate realistic results. The laser beam riding (LBR) guidance system using a semi-active homing quadrant sensor was selected to be the model of choice for the cruise phase, due to the emerging popularity of the system, for military applications such as the FASGW-L project currently active. The LBR missile systems also tend to travel at very high velocities, and require precise controls and robust measurements to detect and track. The systems are the most complex to operate with and will help establish that the proposed imaging system, presented in future chapters, is able to handle the most difficult of scenarios.

Chapter 4

Image Separation

4.1 Introduction

The availability of modern digital cameras has led to the development of a large variety of automated computer vision systems for object identification and tracking. Such systems are always limited, in terms of image resolution, frame rate and field of view (FoV). Ideally, an object tracking vision system should provide both high-resolution imagery for the interrogation of objects within the field-of-view, and a wide field-of-view to maximise the coverage of the environment. Often these two factors are in opposition, with one being compromised to improve the other. Taking inspiration from biology, foveal cameras have been developed [12, 292]. These cameras are based on the human vision system, where the eye has a small, very high-resolution region in the centre of the field-of-view, and a lower resolution peripheral vision. An alternative approach is considered here, where two fields-of-view (one wide, one narrow) are used to generate an image formed from the superposition of the images from the two separate fields-of-view. The resultant image is a superimposed image of the two images, one high-resolution narrow FoV image and one wide FoV with lower resolution. The superimposed image is then separated using a recursive method that is based on the geometry of the two fields-of-view and the spatial relationships between the pixel intensity values as such, it can be considered to be a type of blind image separation, since it does not (in general) require any prior information about the image content.

Separation techniques such as blind source separation (BSS), and in particular independent component analysis (ICA), have been used to solve problems of linear mixtures [56]. BSS is the separation of a set of source signals/data from a set of mixed/observed signals or data. Early work focused on the separation of temporal signals such as audio, hence the common application being the "cocktail party" scenario, where there is a group of people talking simultaneously. Multiple microphones are observing the mixed signals, and the aim is to isolate the source signals from the observed signals. When information about the mixing/convolution of the signals and the source signals are limited, the problem is referred to as "blind". In this scenario, one ap-

proach to estimate the source signals, would be to use some statistical properties from the observed signals to estimate the source signals. It has been shown that, assuming the source signals are statistically independent at any particular instance of time, it is possible to estimate the source signals reasonably well [130]. BSS was recently formulated around 1983 by J. Herault, C. Jutten and B. Ans, but no solution was available at the time [9, 121]. Over the years, the research community steadily grew until the first international workshop solely devoted to BSS was organised in January 1999 [130]. Initially only linear mixtures were considered, but by the end of the 1990s non-linear mixtures were addressed. Furthermore, a general framework for solving BSS problems based on statistical independence was formalised for linear mixtures in 1991 by P. Comon [56, 58], formally referred to as ICA. ICA was originally developed to deal with audio signal processing problems closely associated to the "cocktail party" problem, but since then many further applications have arisen. For example, in biomedical signal processing, such as electrical recordings from an electroencephalogram (EEG), used to measure brain activity. The observed electrical potentials are generally generated by some mixing of the underlying brain components. In certain cases, it is advantageous to understand the original source components of the electrical signals, and from where in the brain the signals originate, to make accurate diagnoses [130]. BSS/ICA can be used for multi-dimensional data, for example in image processing applications. ICA can be used to detect, estimate and remove, noise, blur, clutter or any forms of interference/degradation present in an image, in an attempt to restore an image. It can also be used for feature extraction applications, where suitable representations of an image, audio, or other kinds of data are needed for image registration, data compression or data restoration (i.e. image denoising or removing audio interference) [130]. ICA aims to find the sources signals that are minimally correlated or maximally independent in a probabilistic or information theoretic manner. The problem was first addressed using a neural approach in [121], in which the solution can be viewed as a recursive linear adaptive filter [140]. There are two broad ways in which the independence can be defined for ICA methods, these are: *a)* minimisation of the mutual information, usually measured using entropy estimators such as the Kullback-Leiber Divergence or maximum entropy [130] or *b)* maximisation of non-Gaussianity using kurtosis or negentropy measures [61,131]. Another approach similar to the minimisation of mutual information is the maximum likelihood approach [207]. Farid has offered a potential solution by using principal component analysis (PCA) [82] plus whitening followed by ICA [43, 55]. More recently, Bronstein [33] utilised the sparse ICA (SPICA) technique without any prior knowledge of the image/scene in order to separate a transmitted image from a reflected image. Well-known, ICA algorithms include infomax [24], FastICA [129] and JADE [44, 45], which are publicly available.

Unfortunately, traditional image separation techniques such as PCA and ICA, require at least two sources of "observed" data/signals to perform the separation. However it is important to note that, in this work there is only a single superimposed image frame (one observed signal/image) available at any point in time. As such, it is an under-determined separation problem [57], in this case all BSS methods are unable to provide a technique to separate the superimposed image into its respective $nFoV$ and $wFoV$ images. Another issue likely to be encountered is the equal superposition of the images, as there is no obvious bias towards any of the component images, all the BSS techniques described above are unable to separate the superimposed image. Additionally, BSS methods assume independence and non-Guassianity of the source signals [130]. In this case a high proportion of common information is present because the narrow field-of-view ($nFoV$) image is a more detailed (higher spatial frequency) version of the central region of the wide field-of-view ($wFoV$) image. This combines to cause poor separation using a BSS method, being unable to separate the superimposed image into its constituent parts. Hence, a different approach is needed to solve the problem. In our proposed scenario, the geometric relationship of the superimposed fields-of-view are used to assist the separation process.

4.2 Theory

Envision two images of the same scene, co-boresighted, but with differing fields-of-view, a wide field-of-view image covering a large area of interest, and a narrow field-of-view image magnifying the central region of the $wFoV$. The images have the same numbers of pixels, however the $nFoV$ image will have finer spatial resolution due to the optical magnification of the $nFoV$ onto the sensor. These images are combined - either directly on to the sensor, using two optical paths, or captured and combined using a switching mechanism. The combined superposition (superimposed) image is then recorded and stored as a single digital image. The digital image must then be separated (post-processed) for presentation of the separate images to the end user.

The two source images are aligned to have a common centre and the superposition is an equal addition of the two images, calculated using Equation (4.2). The fields-of-view of the $nFoV$ and $wFoV$ images are fixed and known either as a preset or via metadata attached to the superposition image.

The separation method presented here is based on the (known) geometry of the superposition of the two images. With the two centres of the images aligned, the spatial structure has a radial correlation. One point in the narrow field-of-view is correlated with a point in the wider field-of-view, which is dependent on the relative fields-of-view of the two images.

Only the scaling ratio of the source imagery is required, the ratio γ is defined as:

$$\gamma = wFoV/nFoV \tag{4.1}$$

where $wFoV$ is the wide image field-of-view (degrees or radians or unitless) and $nFoV$ is the narrow image field-of-view, hence γ is always defined to be $\gamma > 1$. Note $\gamma = 1$ is the trivial case where there is only one source image.

We simulate a superimposed image, I_m, using a linear equation:

$$I_m = cI_w + dI_n \tag{4.2}$$

$$c = 0.5 \qquad d = 0.5 \tag{4.3}$$

where c and d are constants representing the intensity contribution of each image. It is assumed that the superimposed image is composed of equal intensities from both images.

Equation (4.4) - (4.7) illustrate how the problem can be formulated. The conventional origin of the image is relocated from the top left corner to the centre of the image for calculations involving the axes α and β (image angles), as shown in (Fig. 4.1). The intensity from the $wFoV$ image is denoted by $I_w(\alpha, \beta)$ and the $nFoV$ image is $I_n(\alpha, \beta)$. The values α, β are transformed into the image matrix coordinate system (X, Y) in order to perform pixel identifications shown in (Fig. 4.1). Equation (4.4) shows that the superimposed image's intensity at a particular pixel location consists of a proportion of the intensity from the $wFoV$ image and a proportion from the $nFoV$ image. Likewise, the subsequent related pixel locations along the diagonal from the outer pixel of interest (4.4) at (α_0, β_0), working inwards towards the centre pixel (4.7).

$$I_m(X_0, Y_0) = c\ I_w(\alpha_0, \beta_0) + d\ I_n(\gamma^{-1}\alpha_0, \gamma^{-1}\beta_0) \tag{4.4}$$

$$I_m(X_1, Y_1) = c\ I_w(\alpha_1, \beta_1) + d\ I_n(\gamma^{-1}\alpha_1, \gamma^{-1}\beta_1) \tag{4.5}$$

$$I_m(X_2, Y_2) = c\ I_w(\alpha_2, \beta_2) + d\ I_n(\gamma^{-1}\alpha_2, \gamma^{-1}\beta_2) \tag{4.6}$$

$$\vdots$$

$$I_m(X_N, Y_N) = c\ I_w(\alpha_N, \beta_N) + d\ I_n(\gamma^{-1}\alpha_N, \gamma^{-1}\beta_N) \tag{4.7}$$

Scaling inwards from the outer pixel (α_0, β_0) obeys the following relation:

$$(\alpha_{k+1}, \beta_{k+1}) = (\gamma^{-1}\alpha_k, \gamma^{-1}\beta_k) \tag{4.8}$$

where $k \to N$, and N is the number of equations in the closed set of "linked pixels" required to reach the centre pixel. This relation can be written as a direct result rather than a sequence, with respect to the outer pixel (α_0, β_0) as:

$$(\alpha_k, \beta_k) = (\gamma^{-k}\alpha_0, \gamma^{-k}\beta_0) \tag{4.9}$$

substituting this into equation 4.4 - 4.7 yields:

$$I_m(X_0, Y_0) = c\ I_w(\alpha_0, \beta_0) + d\ I_n(\gamma^{-1}\alpha_0, \gamma^{-1}\beta_0) \tag{4.10}$$

$$I_m(X_1, Y_1) = c\ I_w(\gamma^{-1}\alpha_0, \gamma^{-1}\beta_0) + d\ I_n(\gamma^{-2}\alpha_0, \gamma^{-2}\beta_0) \tag{4.11}$$

$$I_m(X_2, Y_2) = c\ I_w(\gamma^{-2}\alpha_0, \gamma^{-2}\beta_0) + d\ I_n(\gamma^{-3}\alpha_0, \gamma^{-3}\beta_0) \tag{4.12}$$

$$\vdots$$

$$I_m(X_N, Y_N) = c\ I_w(\gamma^{-N}\alpha_0, \gamma^{-N}\beta_0) + d\ I_n(\gamma^{-(N+1)}\alpha_0, \gamma^{-(N+1)}\beta_0) \tag{4.13}$$

Figure 4.1: Illustration defining X,Y, α and β using standard Lena image

Equations (4.10) - (4.13), define the relationship between the corresponding pixels that link the $nFoV$ scene to the $wFoV$ scene by an iterative process. The distances $(\gamma^{-1}\alpha)$ and $(\gamma^{-1}\beta)$ will tend towards zero. However, as the pixel is the smallest resolvable spatial unit, the iteration is terminated at the centre pixel.

The next step is to identify that the intensities $I_w(\alpha_k, \beta_k)$ and $I_n(\alpha_k, \beta_k)$ are approximately equal because of the high degree of spatial correlation in most images. This means that the equations (4.10) - (4.13) form a closed set, which can be solved recursively, for each pixel intensity value. For example, taking equations (4.10) - (4.11), the linked pixels show that the $nFoV$ intensity value in (4.10), is equal to the $wFoV$ intensity value in (4.11). Similarly, in (4.11) - (4.12), the $nFoV$ intensity value in (4.11), is identical to the $wFoV$ intensity information in (4.12). This identification is made throughout the closed set of equations, such that it can be expressed as:

$$I_n(\gamma^{-k}\alpha_0, \gamma^{-k}\beta_0) = I_w(\gamma^{-k}\alpha_0, \gamma^{-k}\beta_0) \tag{4.14}$$

The set of equations ((4.10) - (4.13)), now only have two unknown intensity values, $I_w(\alpha_0, \beta_0)$ in (4.10) and $I_n(\gamma^{-(N+1)}\alpha_0, \gamma^{-(N+1)}\beta_0)$ in (4.13).

To solve for the intensity value $I_w(\alpha_0, \beta_0)$ in (4.10), we take a reverse count of $l = N - 1$ to 0, where N is the number of "linked" pixels i.e. $\mathbf{D} \rightarrow \mathbf{C} \rightarrow \mathbf{B} \rightarrow \mathbf{A}$ in Fig. 4.2

$$I_w(X_l, Y_l) = \frac{I_m(X_l, Y_l) - d\ I_w(X_{l+1}, Y_{l+1})}{c} \tag{4.15}$$

where the process is initialised at the centre pixel, because the two pixel locations for the $nFoV$ and $wFoV$ are the same, we assume that:

$$I_w(X_N, Y_N) = I_m(X_N, Y_N) \tag{4.16}$$

hence, the $wFoV$ image can be recovered. The closed set of equations is generated for each "outer pixel" location (i.e. **A** in 4.2) and recursively solved as described above, in a pixel-by-pixel fashion, until the entire image is recovered. The $nFoV$ image can then be recovered in one step after the $wFoV$ image has been obtained:

$$I_n = \frac{I_m - c\ I_w}{d} \tag{4.17}$$

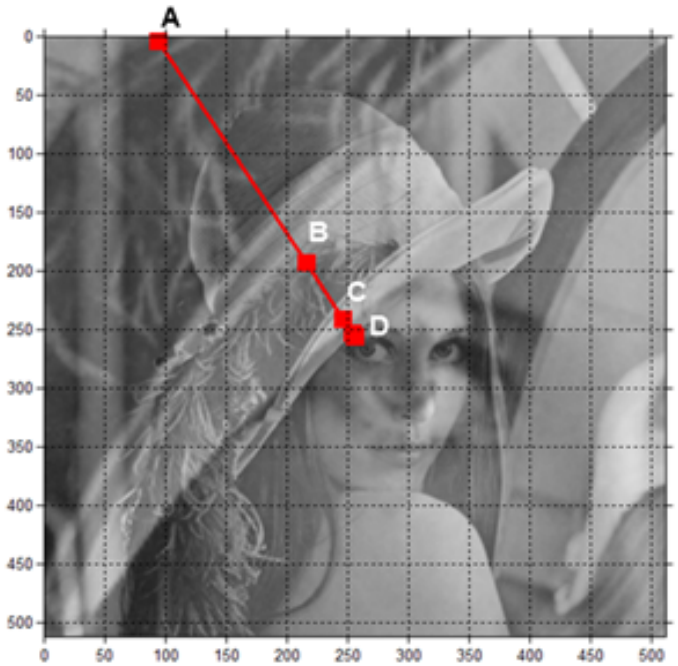


Figure 4.2: Illustration of the related pixels between the two images of the same scene. The furthest pixel away from the centre is the pixel in question and the others are the related pixels which have common information.

Figure 4.2 illustrates the relationship of the current pixel being processed, **"A"**, with its related pixels within the image. Pixel **"A"** represents the point (α_0, β_0) similarly pixel **"B"** represents $(\gamma^{-1}\alpha_0, \gamma^{-1}\beta_0)$ and likewise pixel **"C"** is located at $(\gamma^{-2}\alpha_0, \gamma^{-2}\beta_0)$ in the angular FoV coordinate system as expressed in (4.10) - (4.12). Pixel **"A"** contains a proportion of common information shared with pixel **"B"**, which is from the respective $wFoV$ image, this is evident by comparing equation (4.10) with (4.11). Pixel **"B"**, in turn also has commonality with pixel **"C"** due to that pixel's overlapping information between its $nFoV$ and $wFoV$ images. This recursion continues until the termination condition is satisfied (i.e. centre of the superimposed image is reached, pixel **"D"**).

The images are recovered via a reverse, bottom-up recursive procedure. By following Fig. 4.2, starting from the centre pixel **D**, the recovered pixel intensity value is calculated. The pixel intensity for **D** is subtracted from pixel **C** following (4.15), the result is then subtracted from pixel **B**, again using (4.15), and likewise the result of this is then subtracted from **A**. Once the outer most pixel is reached i.e. (α_0, β_0), equations (4.15) and (4.17) are then applied to calculate the pixel intensities for the corresponding recovered $nFoV$ and $wFoV$ images. The target point (α_0, β_0) is moved to an adjacent pixel and the process is repeated throughout the entire image.

In order to improve the quality of the recovered images, a bilinear interpolation method has been used to obtain improved estimates for the pixel intensity values. Since (α, β) will often represent fractional pixel coordinates, accessing pixel intensities based on a simple rounding to integer coordinates (intra-pixel rounding) can be problematic unwanted quantisation artefacts can be introduced by selections occurring near pixel boundaries. Rounding can cause unwanted artefacts in the separated images due to the potentially extreme transitions between pixel intensities across boundaries, these imprecise selections manifest as a macro-blocking effect - and are most noticeable after a radial line (Fig. 4.5 and 4.6) passes through a region of high contrast. Macro-blocking can be visually interpreted as an undesired mosaicing/tiling-like effect present on images [163]. A comparison of the interpolation method against the non-interpolation method is also presented in the next section.

4.3 Results

The results were generated using an 8-bit greyscale Lena image at a resolution of 512×512 with a $\gamma = 2$ and $\gamma = 4$ in order to compare how well the algorithm works with different ratios between the $nFoV$ and $wFoV$. The widely recognised Lena image was chosen, as it is the most commonly accepted benchmark image within the image processing research community. The superimposed images used for the separation are shown in Fig. 4.3 and Fig. 4.4. The superimposed image is created by taking the central section of the original $wFoV$ image; this creates the $nFoV$ image. The $wFoV$

image is then scaled down to the same size as the $nFoV$ image in order to simulate finer spatial detail in the $nFoV$ in comparison to the $wFoV$ image. The images are then superimposed ($c = d = 0.5$).

The structural similarity (SSIM) index is a method used for measuring similarities between two given images. It is particularly of use as it based on local structural information differences, along with luminance and contrast changes, between the images, in order to measure the degradation [275, 276]. Further details on SSIM and other image quality metrics are provided in Chapter 7. Table 4.1 shows the results obtained when comparing the recovered images to the ideally separated images. The results show that more useful structure is recovered in general when the bilinear interpolation method is used compared to without using the implementation.

γ	Without Bilinear Interpolation		With Bilinear Interpolation	
	$nFoV$	$wFoV$	$nFoV$	$wFoV$
2	0.5279	0.5526	0.8721	0.8588
4	0.7633	0.7947	0.9649	0.9645

Table 4.1: Structural similarity (SSIM) index [276] results for comparing the recovered images against their "ideal" (source) images

Fig. 4.5 **(Top)** presents the $wFoV$ recovered image without the use of the bilinear interpolation. Fig. 4.5 **(Bottom)** shows the respective $nFoV$ recovered image. The macro-blocking artefacts can be seen due to the inaccuracy of accessing the correct related pixels. Macro-blocking is typically associated with lossy signal compression techniques especially with DCT related compression methods such as JPEG, H.261 and certain variants of the MPEG extension [54, 94, 142, 163]. The macro-blocking can be visually interpreted as a undesired mosaicing/tiling-like effect present on images. It is important to note, that in this case the macro-blocking is not due to any form of compression, but the artefacts created due to the intra-pixel boundary problem, reproduces artefacts similar to the JPEG compression technique. By carefully inspecting Fig. 4.5 **(Top)** it can be seen that there is a ghosting effect remnant from the separation process. When the pixel location is too close to a pixel boundary between two neighbouring pixels, the wrong pixel intensity value can be chosen. This is particularly prominent when the wrong pixel is chosen near a large gradient change in intensity i.e. edges. This can be largely seen in the top-left corner of both images in Fig. 4.5 due to the fine spatial structure of the embellishment on the hat, obtained from the $nFoV$ subset of the superimposed image.

Comparing Fig. 4.5 **(Top)** with Fig. 4.6 **(Top)** it can be seen that the $wFoV$ image recovered results in a better separation, with reduced macro-blocking artefacts. The macro-blocking effect is noticeable primarily in the upper-left proportion of the images. This apparent reduction in macro-blocking is due to the overall reduced variation in

intensity in the superimposed image as the relative spatial scaling between the $nFoV$ and $wFoV$ images increases. Observing the recovered $nFoV$ image in Fig. 4.6 **(Bottom)**, the finer spatial details of the embellishment on the hat has been recovered but the macro-blocking is still evident.

Comparing Fig. 4.5 with Fig. 4.7, with $\gamma = 2$, it can be seen that the recovered images using the bilinear interpolated intensities have a much clearer reproduction of the recovered FoV images. It can be seen that the macro-blocking has been reduced significantly and likewise for the ghosting effect along with better contrast reproduction. A similar improvement is found in the case for $\gamma = 4$ in Fig. 4.6 and Fig. 4.8. Both of these improvements are reflected in the SSIM values given in Table 4.1, with the improvement in the case $\gamma = 4$ being marginally better.

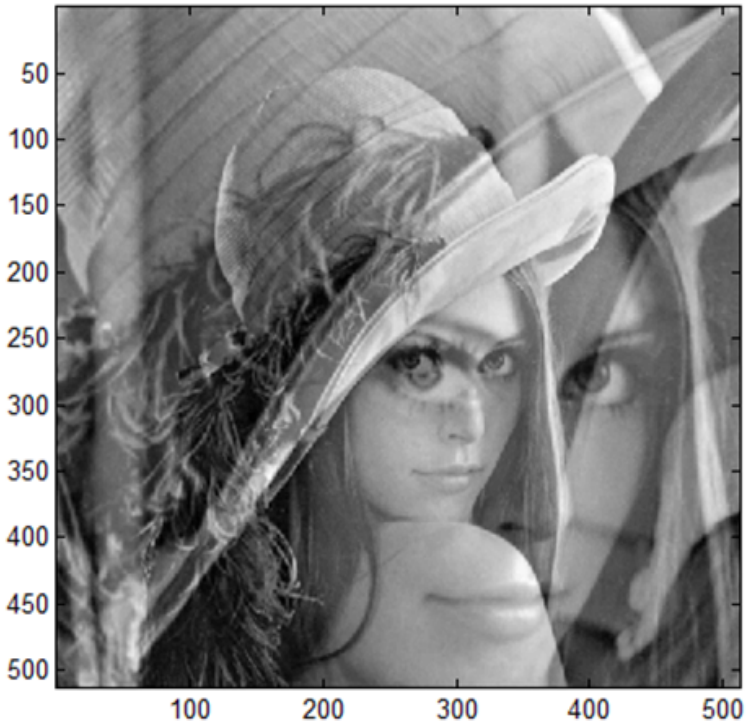


Figure 4.3: Superimposed image using standard Lena image for $\gamma = 2$ at a resolution of 512x512

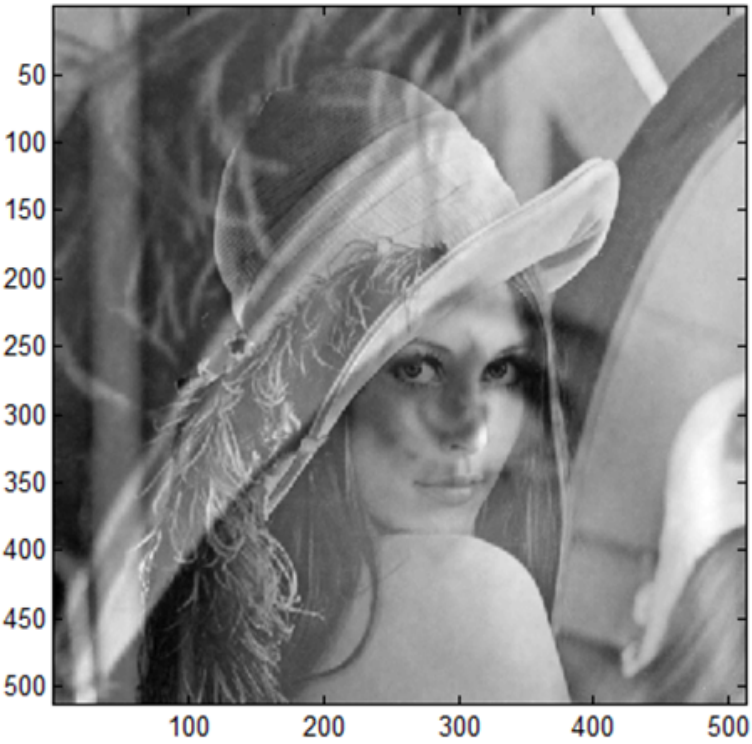


Figure 4.4: Superimposed image using standard Lena image for $\gamma = 4$ at a resolution of 512x512

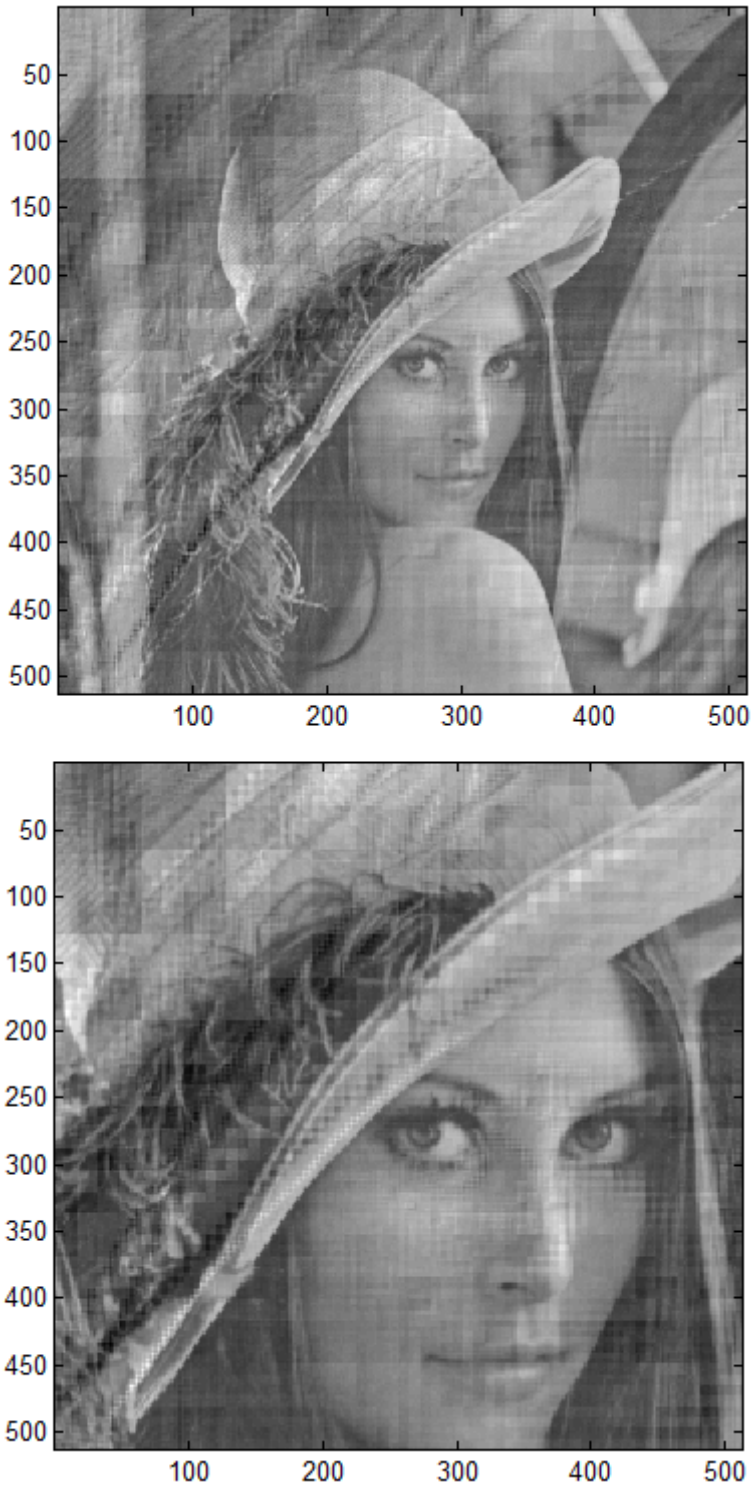


Figure 4.5: Recovered images with $\gamma = 2$ without the use of bilinear interpolation. The above images exhibit macro-blocking due to the quantisation of the pixel locations. **(Top)** *wFoV* recovered image. **(Bottom)** *nFoV* recovered image

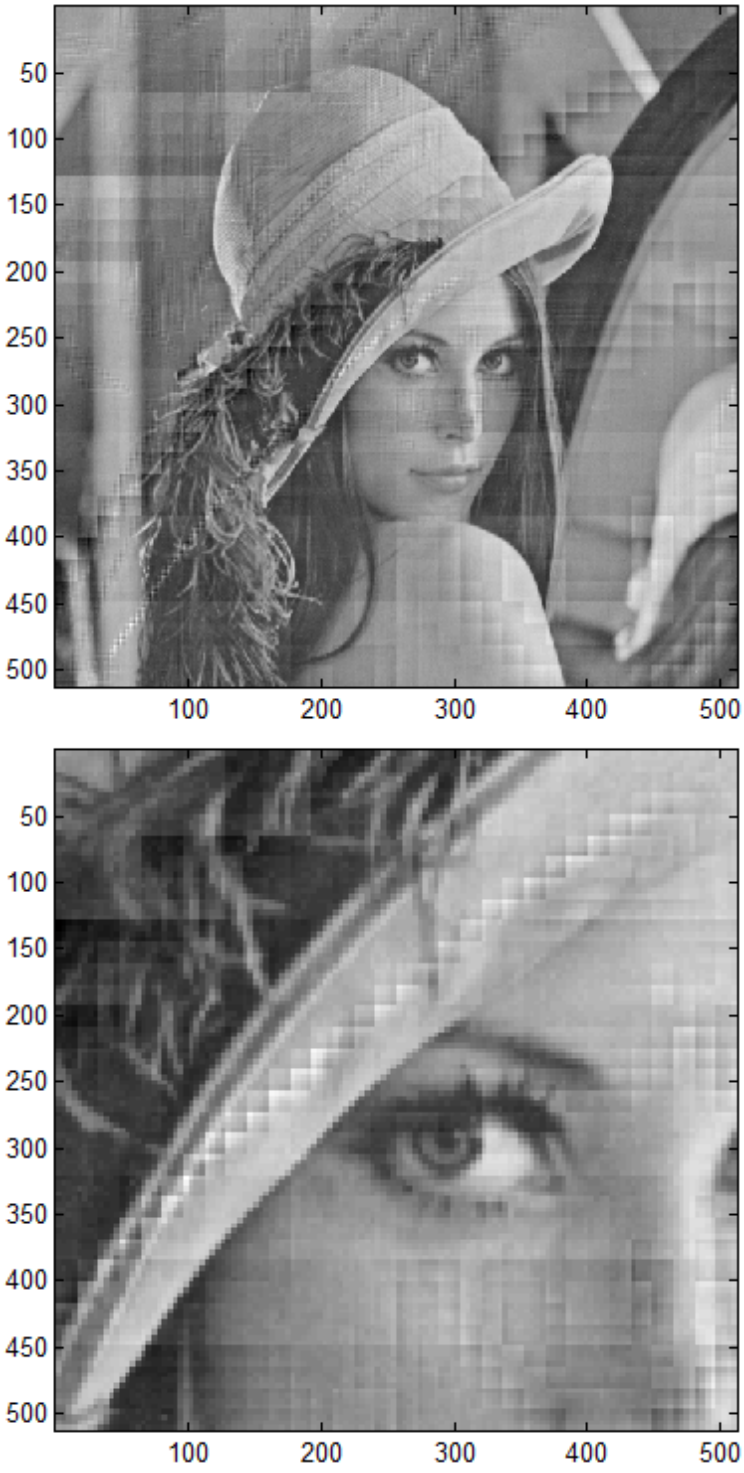


Figure 4.6: Recovered images without using bilinear interpolation with $\gamma = 4$. **(Top)** $wFoV$ recovered image **(Bottom)** $nFoV$ recovered image

Figure 4.7: Recovered images using bilinear interpolation with $\gamma = 2$.
(Top) $wFoV$ recovered image **(Bottom)** $nFoV$ recovered image

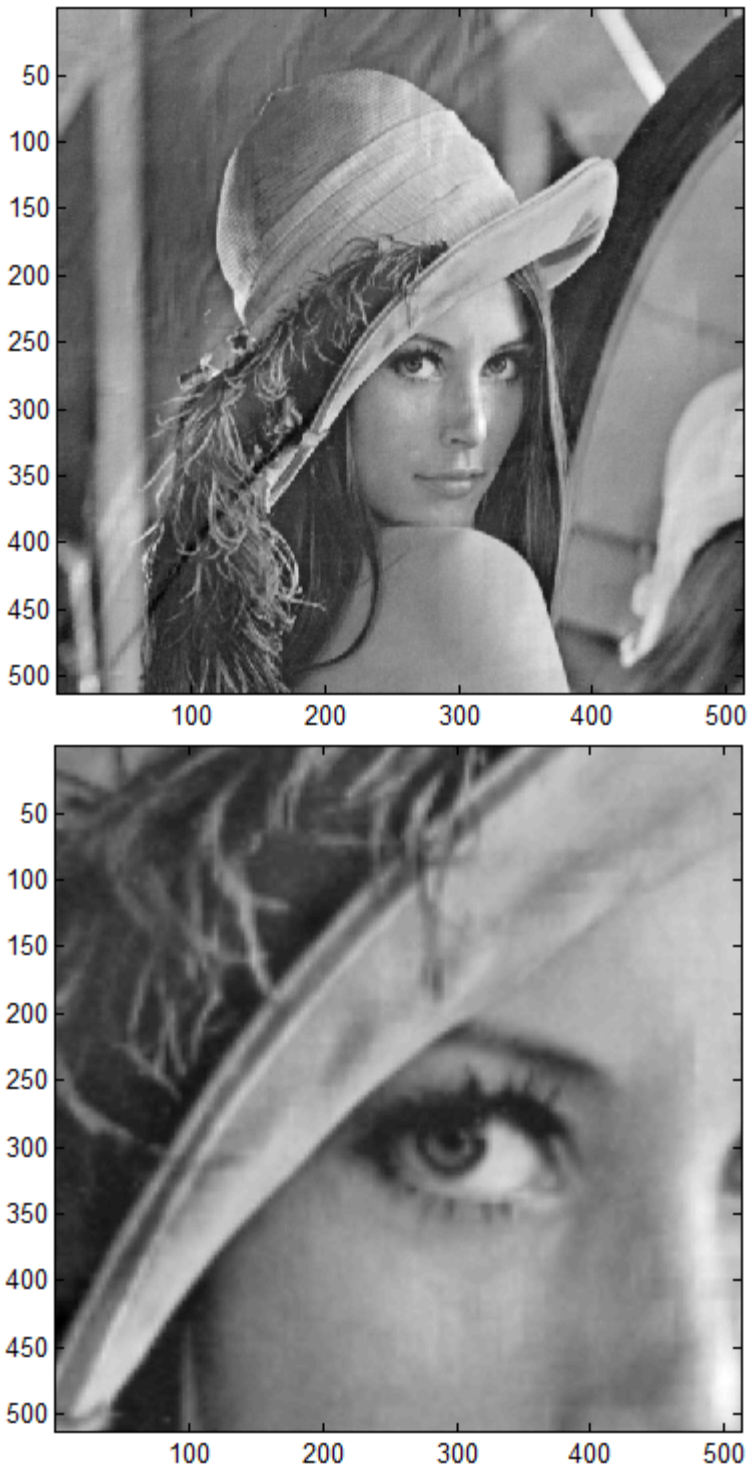


Figure 4.8: Recovered images using bilinear interpolation with $\gamma = 4$.
(Top) $wFoV$ recovered image **(Bottom)** $nFoV$ recovered image

4.4 Chapter Summary

A superimposed image containing two images was generated. Each image was of the same scene but with different fields-of-view. The superimposed image was then separated into the two component (recovered) images. Independent component analysis (ICA) is unable to separate the superimposed image due to the large amount of common information content present between the two images, and along with the images being of equal intensities within the superimposed image.

Fundamentally, this is an under-determined problem i.e. the number of input sensors is less than the number of outputs. However, there is additional information present in terms of the geometric relationship between the two fields of view. The proposed method uses the geometric relationship between the two fields of view to construct a set of recursive equations for the pixel intensities for the two separated images. The recursive equations were solved for an example image (Lena) for two different ratios between the fields of view, $\gamma = 2$ and $\gamma = 4$.

The proposed method was found to separate the images, from the superimposed image effectively, but the intra-pixel rounding of the image angle values was seen to generate block-like artefacts in the resultant images. To mitigate these artefacts, a bilinear interpolation method was used to reduce the effect of the intra-pixel rounding operations. The result of this interpolation was a slight smoothing of the separated images, but this was offset by a significant improvement in the image quality, both visual (qualitative) and in terms of the calculated SSIM metric (quantitative).

Chapter 5

Image Registration

5.1 Introduction

The previous chapter presented and evaluated the geometric image separation technique. This chapter provides the precursory knowledge required to extend the geometric image separation technique to work with multiple frames such as live-streaming videos. Videos can be considered to be a series of image frames sequentially ordered. Most modern cameras can record full high definition (FHD) ($1920px \times 1080px$) videos at 60fps (frames per second) or even higher. Some high-end cameras can record up to quad HD (QHD) ($2560px \times 1440px$) or even 4K ($3840px \times 2160px$) at 60fps. Higher resolution cameras allow to capture better images with higher image detail. Modern high frame rate cameras can record up to 120fps and slow-motion cameras can capture up to $10,000$fps. As the frames rates continue to increase, there is higher likelihood of mutual image data between neighbouring frames, at the same time there is additional information obtained from either a change in the image scene or desired/undesired camera motion. The rapid development of image acquisition technology, has spurred the need for more robust and accurate registration techniques for images with higher quality (i.e. frame size, frame rate, colour and dynamic range) [300].

Image registration is an image processing/computer vision technique that can be used to spatially align and/or overlay multiple images of a captured scene into a single image. One image, usually termed the reference or source image, is kept unchanged and used as the basis for the transformation. The other image(s) are known as the sensed images, which are geometrically transformed to be spatially aligned/overlaid with the reference image. Image registration aims to find one or more spatial transformations, so that two or more images of the same scene can be geometrically transformed and spatially aligned [106]. The images can be of the same scene but captured from a different viewpoint, at a different time and even from a different imaging sensor [34]. The applications span over many research fields not limited to remote sensing [105,190], video compression [152] and medical imaging [300].

The registration techniques can be divided into four main application groups, de-

pending on the variety of different acquisition methods, and the final resultant image. The four main application groups are: multiview, multitemporal, multimodal and scene to model registration [26, 300].

Multiview i.e. Input images captured from different viewpoints, in order to obtain a new perspective of the same scene for better understanding of the scene either in 2D or 3D. For example, the 3D reconstruction of the human brain using MRI (magnetic resonance imaging) or CT imaging devices [6] or for remote sensing applications such as mosaicing of images of a surveyed area [100, 231].

Multitemporal i.e when images of the same scene are acquired at different times, often on a regular basis and often under different conditions [300]. A possible application would be to evaluate changes in a scene over time. Examples would include the monitoring of landscape changes in remote sensing applications or the monitoring of healing of a patient in medical applications i.e. x-ray of broken bones.

Multimodal where images of the same scene are captured by different imaging sensors, allowing the integration of information between different source streams by different sensors. Examples include image fusion for medical applications of the same anatomical body area using different sensors such as MRI, ultrasound, CT/CAT, for better diagnosis and treatments.

Scene to model registration is where images of a scene and a virtual model are registered, so that the images can be used for comparison. Applications include image registration of aerial satellite imagery to incorporate into GIS (geographic information system) maps [5, 111, 126, 143]. See Figure 5.1.

Multimodality imaging and image mosaicing [35,136] are two examples, where image registration tools are crucial. In medical imaging applications [4, 122], multimodality imaging is the process of overlapping images of the same subject/scene from different modalities, such as magnetic resonance imaging (MRI), optical imaging, electroencephalography (EEG) and x-ray imaging [291]. Multimodality imaging allows the observer to have the exact spatial relationship between the various modalities, which can be a necessity in certain situations [19, 79, 282].

Image mosaicing can be a very powerful tool for unmanned aircraft system (UAS) operations. Mosaicing is the process of stitching multiple overlapping images together producing a single large high resolution image. In UAS operations, mosaicing allows the generation of a large scene, which could not be possible using a single conventional camera system and can be very useful for photographers, geographical surveying and cartographers. Recent technological developments has lead to the autonomous real-time ground ubiquitous surveillance imaging system (ARGUS-IS), jointly developed

Figure 5.1: Example of Scene to model registration used by Google Maps©. Images are obtained from multiple source (i.e. DigitalGlobe [1], US Navy [37], Zenrin [103], AutoNavi [153]) and registered to create google's own geospatial GIS maps and then to create 3D models based on scene registration [181].

by DARPA© and BAE Systems©. The ARGUS system is an array of 368 smaller 5megapixel smartphone camera sensors used to create extremely large 1.8gigapixel resolution wide field-of-view mosaic images at 12fps for video surveillance with the ability to track up to 65 simultaneous specified targets. The imaging system has been built to be deployed on a new UAS developed by Boeing© called "Hummingbird" [10, 23].

A broad range of techniques for image registration have been proposed for a variety of applications [34, 300]. There is no comprehensive method capable of all registration tasks due to the various acquisition methods and degradation types. Existing image registration techniques can be classified into two main categories: **area** based and **feature** based methods [81]. The differences between both registration types in sections 5.4 and 5.5 respectively, but in general, the majority of image registration techniques follow four main steps:

Feature Detection is the process of detecting features/areas of an image, such as distinctive objects i.e. lines, edges, corners) or salient regions. These detected features are known as control points and can be either manually or preferably automatically detected. Ideally the control points should be detectable over all possible images of the same scene and invariant to any geometric deformations (translations, rotations, scale) or photometric variations (i.e. brightness, contrast).

Feature Matching locates the corresponding matching feature points between the reference image and other images of the scene. The matching process is based

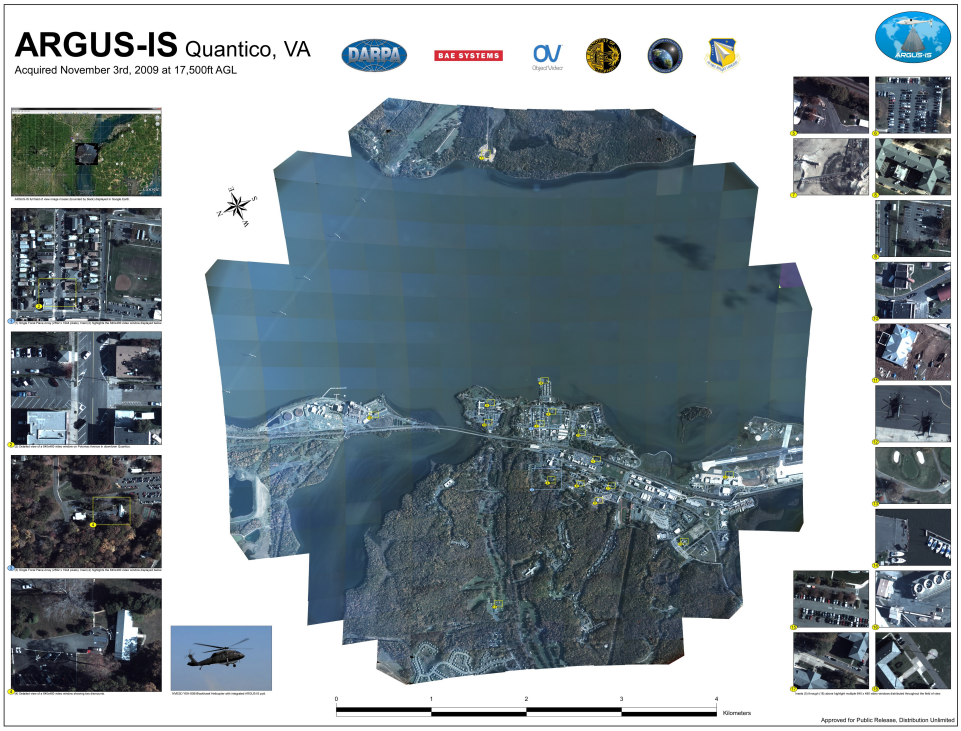


Figure 5.2: ARGUS-IS example of image mosaicing used in ARGUS-IS UAS [10].

on similarity measures of the descriptors constructed from the detected feature points. To facilitate the matching process, the descriptor ideally should be distinctive and insensitive to local image deformations [263]

Transformation Estimation. At this stage, the spatial transformation required to geometrically align two images is estimated. The transformation can be estimated from the matched feature pairs between the images.

Transformation & Image resampling. Once the spatial transformation has been estimated, the sensed image can be aligned/mapped with accordance to the spatial ordinates of the reference image. The registered image is likely to have non-integer pixel co-ordinates generated from the transformation process. To alleviate this problem, the sensed image is generally resampled. The selection of an interpolation technique, can vary depending on the severity of the alignment errors. Typically bilinear interpolation is sufficient, but the interpolation method can range from basic linear interpolation to pyramid blending.

5.2 Transformation Model

The transformation stage is a two step process, first a geometric transformation that modifies the pixel locations of the sensed image to the reference image, secondly a photometric transformation that calibrates the intensity values between the images [105].

A geometric transformation can be mathematically described using an appropriate transformation model to relate the spatial coordinates between two images. The transformation process is a key process in the image registration process. In this section, the images are assumed to differ only by geometric transformation. The techniques for photometric transformations will be examined separately.

A number of possible transformation models exist that can describe the differences between two or more images. For simplicity, it will be assumed that the transformations that relate the two images are limited to the fundamental set of 2D planar transformations: translation, Euclidean (translation, rotation), similarity, affine (scaling, shear) and projective, discussed in detail in Section 5.3. The images I_R (reference) and I_S (sensed) are assumed to have been obtained using an idealised pinhole camera [294, 295].

5.2.1 Pinhole Camera Model

The camera obscura (meaning "dark chamber"), more commonly known as the pinhole camera, is the simplest optical imaging device/model available. It assumes that the camera aperture is infinitely small (i.e. a point) and no lenses are used to focus the incoming light. In its earliest forms, it was the shape of closed box or chamber with a small hole in one of its sides which via rectilinear propagation of light creates an image of the outside space on the opposite side of the box.

One of the earliest depictions of the pinhole camera was published in the book *De radio astronomico et geometrico liber* [96] in 1545. Dutch physicist and mathematician Gemma Frisius described his observation of the solar eclipse in 1544. Johannes Kepler observed sun spots in the same way in Prague at the beginning of the 17th century; Kepler was also the first to use the term camera obscura [14, 220, 221].

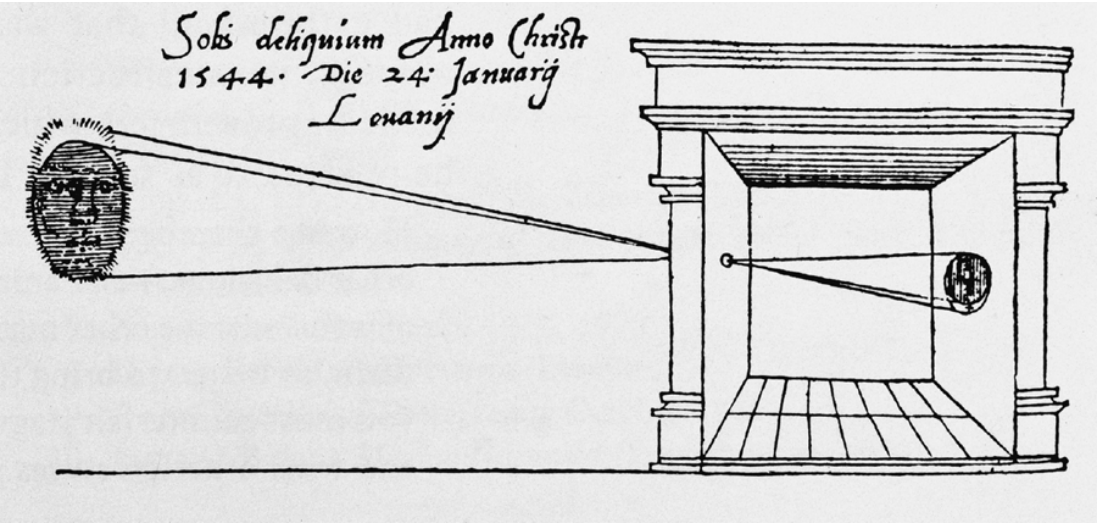


Figure 5.3: First depiction of camera obscura [96]

Figure 5.4 presents a 3D illustration of the basic setup. Point $P(-x_1, x_2, x_3)$ and point $Q(y_1, -y_2)$ represents the corresponding projected point, onto the 2D image/focal plane. In the 3D coordinate system (X_1, X_2, X_3) the origin is at **O**, which is also where the camera aperture is located, or in the case of the pinhole camera, the pinhole through which the light enters the dark chamber [169].

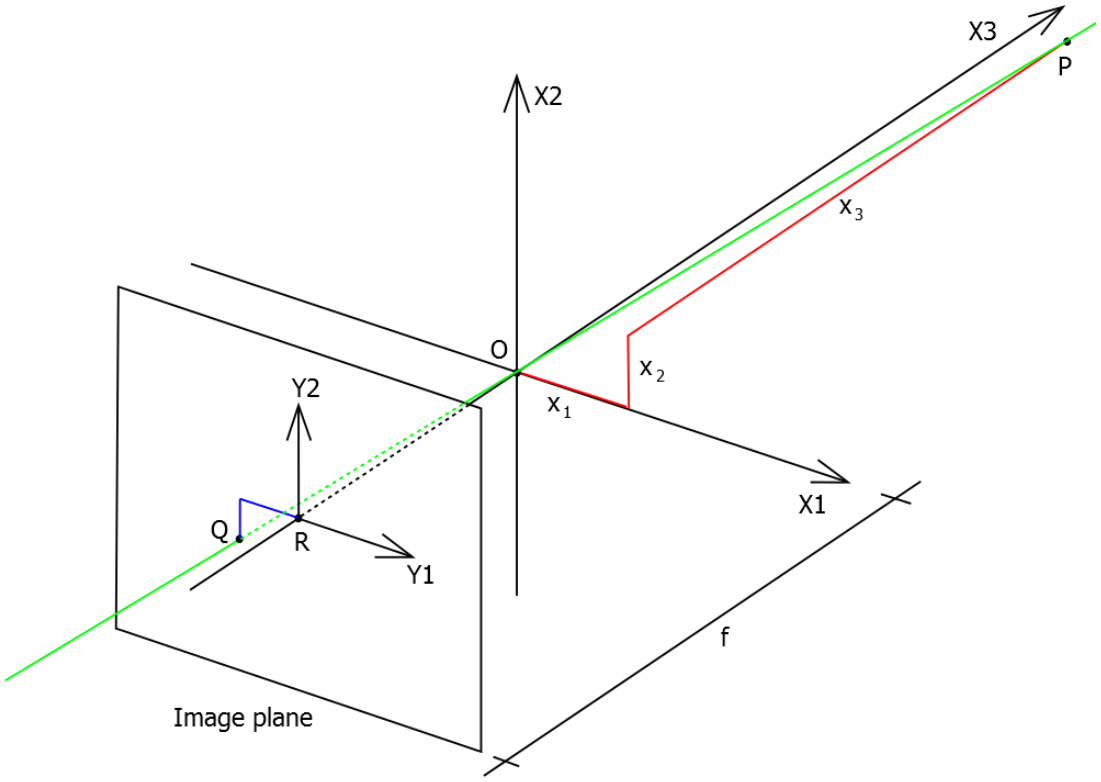


Figure 5.4: Illustration of pinhole geometric model

Axis X_3 is pointing in the viewing direction of the camera and is referred to as the optical/principle axis. The 3D (X_1X_2) plane, with the origin at its centre, is the front side of the camera and is referred to as the principle plane. The (Y_1Y_2) plane with the centre, R, is where the 3D world is projected, through the camera aperture and is called the image plane. Point R is located at the intersection between the optical axis and the image plane, referred to as the principle point or the image centre.

Point $P(-x_1, x_2, x_3)$ is an arbitrary point in the world relative to (X_1, X_2, X_3) axes, the green line represents the projection line which travels through point P and O to the image plane, at point Q.

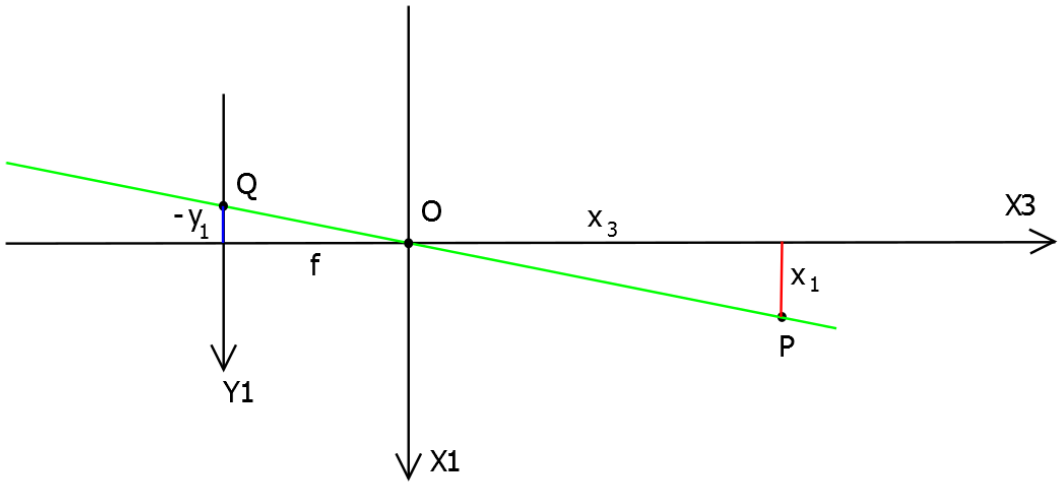


Figure 5.5: Top view (X_2 axis) illustration of pinhole model

From the top view as shown in Figure 5.5, two related triangles come into perspective, with the projection (green) line as there hypotenuses. From geometry, due to the relation of the triangles, it follows:

$$\frac{-y_1}{f} = \frac{x_1}{x_3} \tag{5.1}$$

rearranging to give:

$$y_1 = -\frac{fx_1}{x_3} \tag{5.2}$$

A similar relationship can be deduced by looking from the X_1 axis (i.e. X_2X_3 plane):

$$\frac{-y_2}{f} = \frac{x_2}{x_3} \tag{5.3}$$

rearranging to give:

$$y_2 = -\frac{fx_2}{x_3} \tag{5.4}$$

where f is the focal length of the camera i.e. the distance between the origin and the image plane or in a pinhole camera the distance between the camera aperture and the photographic film or an imaging sensor in a modern digital camera. This can be summarised into matrix form [116]:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{f}{x_3} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{5.5}$$

In a homogeneous (4-dimensional) coordinate system, the pinhole camera model can be expressed as:

$$\lambda \boldsymbol{Y}_{2D} = \boldsymbol{P}\boldsymbol{X}_{3D} \tag{5.6}$$

where $\boldsymbol{Y}_{2D} = [y_1, y_2, 1]^T$ and $\boldsymbol{X}_{3D} = [x_1, x_2, x_3, 1]^T$ are the expressions in the homogeneous coordinates system. $\lambda = -x_3/f$ is the scale factor and $\boldsymbol{P} = [\boldsymbol{I}_{3\times3}, \boldsymbol{0}_{3\times1}]$ is a 3×4

matrix representing the projection relation between $\boldsymbol{Y}_{2D}$ and $\boldsymbol{X}_{3D}$. $\boldsymbol{I}_{3\times3}$ is the 3×3 identity matrix and $\mathbf{0}_3$ a zero 3-vector [116]. The homogeneous 4-dimensional system allows for different centre of projection and different projective coordinate frames in the image. The matrix $\boldsymbol{P}$ is known as the camera matrix. In summary, the action of a projective camera on a point in space can be expressed in terms of a linear mapping of the homogeneous coordinate system [84, 298].

5.2.2 Camera Calibration

The goal of camera calibration is to develop a mathematical model of the transformation between world points and observed image points resulting from the image formation process [211]. The pinhole camera model described in the previous sections is an ideal scenario, and does not take into account intrinsic or extrinsic parameters that are due to camera calibration [84, 93, 298]. These parameters exist and can be accounted for in the camera matrix $\boldsymbol{P}$. The parameters that affect the mapping of world to image points can be separated into three categories:

Intrinsic Parameters

In the ideal pinhole camera, it is assumed that no lens is present in the camera system. In modern cameras, lenses are present and therefore the effect of radial lens distortion must be taken into account. There are three main different types of parameters that would result from an internal calibration. The first type of parameters isα and β, which are the scaling factors in the respective Y_1 and Y_2 directions of the image plane (i.e. the aspect ratio of the image plane may not be unity). The second type is (u_0,v_0), which is the offset coordinate of the principal point (i.e. the image centre) which is the point where the optic axis intersects the image plane, since the principal point may not be at the origin of the image plane (usually the top left corner). The last one is the skew descriptor γ, which represents the skew information of Y_1 and Y_2 axes of the image plane as the image plane may not be perfectly rectangular. The above listed parameters can be summarised into a 3×3 matrix, i.e. the camera intrinsic parameter, termed the camera intrinsic parameter $\boldsymbol{K}_{cam}$:

$$\boldsymbol{K}_{cam} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.7}$$

Distortion Parameters

Distortion parameters describe the geometric non-linearities of the camera [211]. Lens distortion is the most common form of distortion in modern camera systems with lenses. By taking this into account, we can improve upon the original pinhole camera model.

Lens distortion is the situation where an image's pixel locations are displaced from the predicted pixel locations determined by the ideal pinhole model. The most common form of lens distortion is radial distortion where the displacement towards or away from the image centre is proportional to the radial distance [250,284]. Under the assumption that the principal point is the same as the centre of distortion, the correction of radial distortion can be achieved using [298]:

$$\hat{y}_1 = y_1 + L_D(r)y_1 \tag{5.8}$$

$$\hat{y}_2 = y_2 + L_D(r)y_2 \tag{5.9}$$

$(\hat{y}_1, \hat{y}_2)$ are the corrected pixel locations, calibrated to negate lens distortion from the captured image. $L_D(r)$ is known as the distortion function and can be approximated using low-order polynomials:

$$L_D(r) \approx l_{d1}r^2 + l_{d2}r^2 \tag{5.10}$$

l_{d1} and l_{d2} are the radial distortion parameters [73, 138, 272] and the radial distance, $r = \sqrt{y_1^2 + y_2^2}$.

Extrinsic Parameters

Extrinsic Parameters describe the relationship between the camera frame and the world frame, including positions (3 parameters) and orientations (3 parameters) [211]. It aims to relate the 3D scene in the world coordinate system to the $2D$ image plane of the camera coordinate system. The position and orientation parameters considered are the rotation angles θ_1, θ_2, θ_3 around their respective coordinate axes X_1, X_2, X_3. Each angle results in a 3×3 rotation matrix $\boldsymbol{R_1}$, $\boldsymbol{R_2}$, $\boldsymbol{R_3}$ expressed as:

$$\boldsymbol{R_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{bmatrix} \tag{5.11}$$

$$\boldsymbol{R_2} = \begin{bmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix} \tag{5.12}$$

$$\boldsymbol{R_3} = \begin{bmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.13}$$

Additionally there are three translations t_1, t_2, t_3 along three coordinate axes, which are expressed as 3×1 vector $\boldsymbol{T} = [t_1, t_2, t_3]^T$. These rotations and translations can be summarised into a 3×4 matrix, known as the camera extrinsic parameter $[\boldsymbol{R}\ \boldsymbol{T}]$ [93, 298].

$$[\boldsymbol{R}\ \boldsymbol{T}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{5.14}$$

Ultimately, the camera intrinsic and extrinsic matrices can be simplified into the projection matrix $\boldsymbol{P}$, and expressed as:

$$\boldsymbol{P} = \boldsymbol{K_{cam}}[\boldsymbol{R}\ \boldsymbol{T}] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \tag{5.15}$$

5.3 Geometric Transformation Model

5.3.1 Translation

Two-dimensional (2D) translation is the most fundamental geometric transformation model. The 2D spatial relationship functions f_x and f_y can be expressed mathematically as:

$$x' = x + t_x \tag{5.16}$$

and

$$y' = y + t_y \tag{5.17}$$

where t_x and t_y are the translations along the x and y directions respectively. The corresponding homography matrix $\boldsymbol{H}$ has two degrees of freedom and is:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5.18}$$

Homography in its simplest explanation is a way to represent a 2D point (x, y) in an image as a 3D vector $\boldsymbol{x} = (x_1, x_2, x_3)$ where $x = \frac{x_1}{x_3}$ and $y = \frac{x_2}{x_3}$. This is a homogeneous representation of a point which lies on the projective plane, P^2. Homography is an invertible mapping of points and lines on the projective plane, P^2 [74, 116].

5.3.2 Euclidean Transform

A Euclidean transformation consists of a rotation and a translation. It can also be called rigid transformation because of its property of preserve all euclidean distances [300]. The 2D spatial relationship functions f_x and f_y can be written as [250]:

$$x' = (x\cos\theta - y\sin\theta) + t_x \tag{5.19}$$

and

$$y' = (x \sin\theta + y \cos\theta) + t_y \tag{5.20}$$

θ is the rotation angle about the origin. The homography matrix $\boldsymbol{H}$ has three degrees of freedom, and is shown below:

$$\boldsymbol{H} = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5.21}$$

5.3.3 Similarity Transform

A similarity transformation differs from Euclidean transform in that addition to translation and rotation, the similarity transform allow for uniform scaling. The similarity transform preserves angles between lines and hence is often known as a shape preserving mapping. The 2D spatial relationship functions f_x, f_y, can be defined as:

$$x' = s(x \cos\theta - y \sin\theta) + t_x \tag{5.22}$$

and

$$y' = s(x \sin\theta + y \cos\theta) + t_y \tag{5.23}$$

where s is an arbitrary scale factor. The homography matrix has four degrees of freedom, and is defined as:

$$\boldsymbol{H} = \begin{bmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5.24}$$

5.3.4 Affine Transform

The affine transform allows for non-uniform scaling unlike the similarity transform. Non-uniform scaling allows the aspect ratio of the object or image to be altered as well as a shear mapping. Parallel lines stay parallel under the affine transform [250]. The non-uniform scale change can be written as:

$$\begin{aligned} x' &= s_x x \\ y' &= s_y y \end{aligned} \tag{5.25}$$

where s_x and s_y are the scaling values along the x and y directions respectively. The shear mapping can be written as:

$$\begin{aligned} x' &= x + ay \\ y' &= bx + y \end{aligned} \tag{5.26}$$

a and b are the shear factors in the respective x and y directions that cause the pixel distortions. The affine transforms are able to maintain straightness and parallel lines. The 2D spatial relationship f_x and f_y are:

$$x' = a_{11}x + a_{12}y + t_x \tag{5.27}$$

and

$$y' = a_{21}x + a_{22}y + t_y \tag{5.28}$$

The values a_{11}, a_{12}, a_{21}, a_{22} are the affine transformation parameters which factor in rotational, (non-uniform) scaling and shearing components [106]. t_x and t_y are the translational values in the x and y direction, together forming the homography matrix $\boldsymbol{H}$ (which has six degrees of freedom):

$$\boldsymbol{H} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5.29}$$

5.3.5 Projective

The projective transformation model, describes the true imaging geometry [106]. Projective transform employs the homography matrix $\boldsymbol{H}$ directly. It is a non-linear transformation and only preserves straight lines to stay straight [106, 250]. The 2D spatial relationship functions f_x and f_y can be described as:

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + 1} \tag{5.30}$$

and

$$y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + 1} \tag{5.31}$$

The projective transformation's homography matrix, $\boldsymbol{H}$, has eight degrees of freedom and is expressed below:

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \tag{5.32}$$

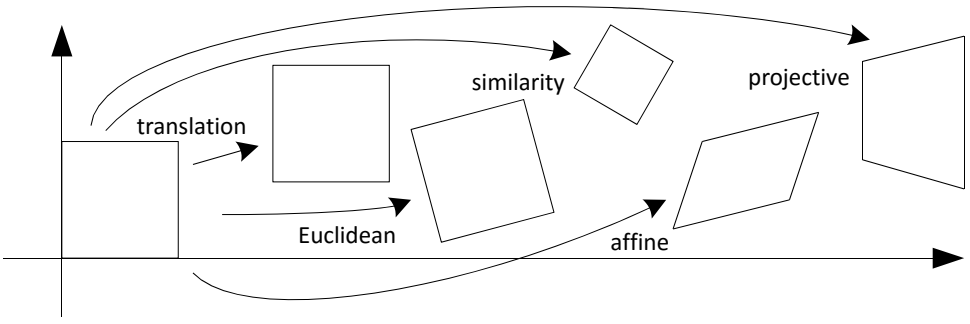


Figure 5.6: Figure illustrating the basic set of 2D planar transformations [250]

Figure 5.6 illustrates the different transformation models and the capabilities of each transformation model.

5.4 Area Based Methods

Area based methods, sometimes called correlation-like methods or template matching [90], favour the feature matching stage rather than the feature detection stage as typically these methods deal with a window of predefined size of an image or the entire image, without the need of detecting salient objects, hence feature detection is effectively merged with the second registration step [210]. Some area based methods also merge the second (feature matching) and third (transform estimation) registration stages by simultaneously searching for feature correspondences and estimating the parameters of the required mapping [137]. Area based techniques combine the direct comparison of the pixel intensity values of two images with optimisation techniques to estimate the transformation required to align/overlay the reference and sensed image(s). The matching step is performed by using a suitable similarity measure function, the sensed image/features are warped relative to the reference image, until the resultant error metric is minimised. The most apparent approach would be to conduct an exhaustive search over the entire image and to try all possible transformations until the optimal transform is established. Below are a few examples of area based registration techniques.

Cross Correlation

Cross correlation (CC) is the most basic statistical approach to registration [34], and is commonly used for template matching or pattern recognition. CC is widely used in other fields such as signal processing or Econometrics [41] and is often called the sliding dot product. In image processing, cross correlation assess the differences in image intensity values over a predefined sized windows [222], or over the entire images, in order to maximise the alignment of the two images [250] without any structural analysis [18]. The CC is able to exactly align mutually translated images only, it can also be successfully applied when slight rotations and scaling are present [300]. The cross correlation value is found to be maximum when the images are exactly matched [224]. The CC of the reference image $I_R(x, y)$ and the sensed image $I_S(x, y)$ that has been translated by (t_x, t_y) can be defined as [48, 251]:

$$CC(t_x, t_y) = \sum_x \sum_y I_R(x, y) I_S(x + t_x, y + t_y) \tag{5.33}$$

Classical area based methods such as CC, exploit direct exploitation of image intensities. Consequently they are sensitive to the intensity changes introduced by noise,

varying illumination and even different sensor types [18]. Due to this the normalised cross correlation (NCC) is often used instead and is expressed as:

$$NCC(t_x, t_y) = \frac{\sum_x \sum_y (I_R(x,y) - \mu_R)(I_S(x+t_x, y+t_y) - \mu_S)}{\sqrt{\sum_x \sum_y (I_R(x,y) - \mu_R)^2 (I_S(x+t_x, y+t_y) - \mu_S)^2}} \tag{5.34}$$

μ_R and μ_S are the average intensity values of the reference and sensed images respectively, and can be defined as:

$$\mu_R = \frac{1}{N} \sum_x \sum_y I_R(x,y) \tag{5.35}$$

$$\mu_S = \frac{1}{N} \sum_x \sum_y I_S(x+t_x, y+t_y) \tag{5.36}$$

where N is the total number of pixels in the image or the window/image patch. The NCC value is always in the range $[-1, 1]$, which makes it easier to handle for higher level applications, such as deciding whether the images truly match. Example results for the NCC, have been presented in Figure 5.8. The NCC tends to degrade in performance for noisy low-contrast regions [250].

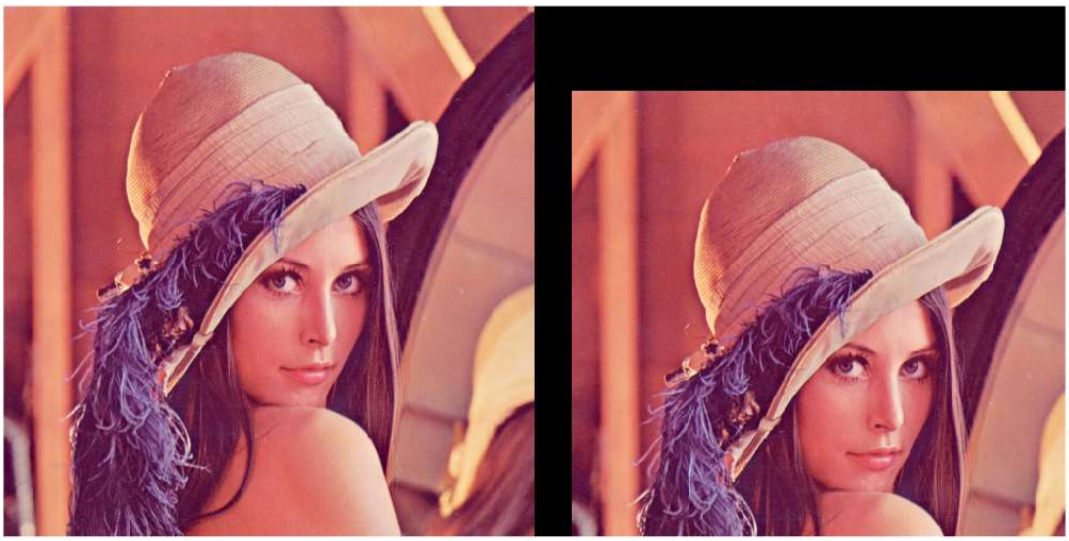


Figure 5.7: Image showing original Lena image (left) and translated Lena image (right) by $[35, 82]px$

The generalised versions of CC can handle more complex geometric transformations, by calculating the assumed geometric transformations of the sensed image/window [114] are able to handle more complex geometric deformations [25, 242]. However, the ability to handle higher complexity deformations, comes at the cost of higher computational load [300]. To address this problem, the Fourier transform transform is used. This is based on the cross-correlation theorem, which states that the CC of two images in the

spatial domain corresponds to the product of the images in the Fourier domain, where one of the terms is the complex conjugate [201]. The CC of the images can be computed more efficiently with the use of the Fast Fourier Transform (FFT). Although using the Fourier domain simplifies the computational processing there are some limitations, firstly the use of the FFT is only applicable to CC and not for NCC. Secondly the computational memory capacity required grows with the logarithm of the image area [34].

To address the limitations of memory capacity and computational processing time, a windowed variant of the CC methods in the Fourier domain is preferable. Instead of using the entire image, a window of predefined size is employed, and is perfectly suitable for situations where the images overlap by a small amount or one image is a small subset of the other image [251].

$$WCC(t_x, t_y) = \sum_x \sum_y w_R(x,y) I_R(x,y) w_S(x+t_x, y+t_y) I_S(x+t_x, y+t_y) \quad (5.37)$$

Equation 5.37 above represents the mathematical expression in the spatial domain, for the windowed cross correlation of the reference image $I_R(x,y)$ and the sensed image $I_S(x,y)$ at a translation of (t_x, t_y), where w_R and w_S are the windowing functions which are zero padded so that the circular shifts return zero values outside of the original images boundaries [251].

As with the previous CC and NCC, the WCC can be adapted to function over higher degrees of freedom, therefore it is not limited to translational deformations. This is done by applying a change of interest in translation, rotation and scale to the sensed image, hence it is able to detect changes in higher degrees of freedom, but with higher computational costs [250].

Sequential Similarity Detection Algorithm

Sequential similarity detection algorithm (SSDA) proposed by Barnea and Silverman [15] is considered to be a more efficient method than the traditional CC methods. There are two major improvements over traditional CC methods. First, they suggest a similarity measure, which is computationally simpler, based on the absolute differences (1-norm distance) between the pixels in the two images [34, 48]:

$$SSDA(t_x, t_y) = \sum_x \sum_y ||I_R(x,y) - I_S(x+t_x, y+t_y)||_1 \quad (5.38)$$

Secondly, SSDA introduces a sequential search strategy. For any translation of the sensed image, if the accumulated sum of the similarity measure (Equation 5.38) exceeds the predefined threshold, it is rejected and the next translation candidate is examined, thus reducing computational complexity. It is important to note, that even so, as

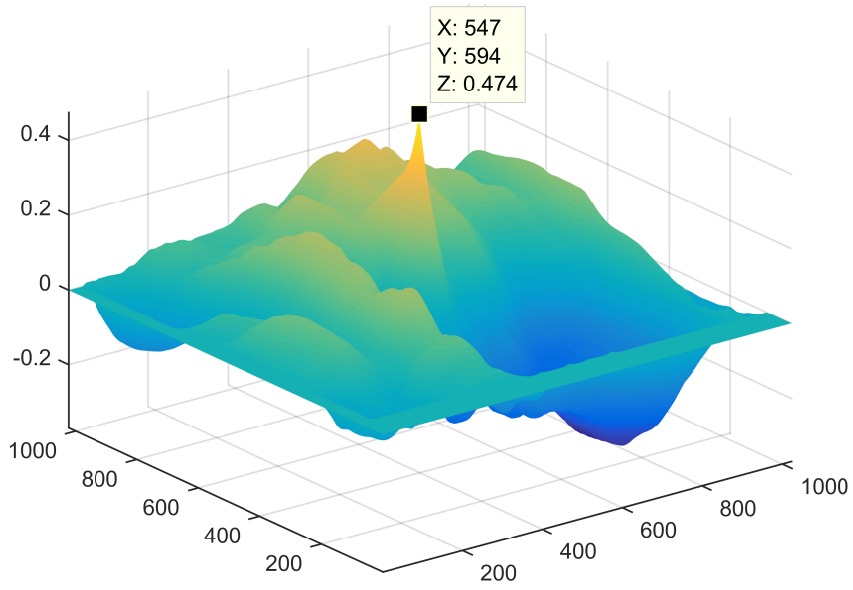


Figure 5.8: Normalised Cross-correlation results of the two Lena images shown in Figure 5.7. The peak value describes the offset in translation (i.e. $[547-512, 594-512] = [35, 82]px$).

the degrees of freedom of the deformations increase, the computational complexity increases.

Both SSDA and CC are unable to deal with dissimilar images, the similarity measures described above (i.e CC and SSDA) are maximised/minimised to find identical matches. For this reason, feature based techniques (discussed in Section 5.5) which measure invariant properties of the Fourier transform are preferable when images are acquired under different circumstances (i.e. weather conditions, lighting, different sensors) [34].

Phase Correlation

To further address the issues of high computational complexity and registration accuracy, the frequency domain offer certain advantages, such as the transform can be efficiently implemented in either hardware or using the FFT [34]. For higher degrees of freedom correlation-like and sequential methods tend to become inefficient, unlike frequency domain methods and are more robust under varying lighting and atmospheric conditions (i.e correlated and frequency dependent noise).

The phase correlation (PC) method by Kuglin and Hines [149] was proposed to estimate the shift between two images based on the shift property of the Fourier transform referred to as the shift theorem [34,300] and is applied to all FFT based methods. The Fourier registration method attempts to determine the location of the peak in the spatial domain of the inverse Fourier transform in the cross-power spectrum phase of

the two images, thus determining the translation difference. The cross power spectrum, also known as the cross spectral density is the Fourier transform of the cross correlation function. Assuming there is only a translational deformation between two images i.e. the reference image I_R and the sensed image I_S described as:

$$I_R(x, y) = I_S(x + t_x, y + t_y) \tag{5.39}$$

The corresponding 2D Fourier transforms of both sides are related by:

$$\mathcal{I}_R(u, v) = \mathcal{I}_S(u, v)e^{2\pi j(ut_x+vt_y)} \tag{5.40}$$

$\mathcal{I}_R(u, v)$ and $\mathcal{I}_S(u, v)$ are the 2D Fourier transforms of $I_R(x, y)$ and $I_S(u, v)$ respectively and (u, v) are the corresponding (x, y) locations in the frequency domain. Each point in the frequency domain (u, v) represents a particular frequency value contained within the spatial domain image [87]. A translational shift in the spatial domain is represented by a phase change in the frequency domain [216]. The normalised cross power spectrum can be calculated as:

$$\begin{aligned} \mathcal{PC}(u, v) &= \frac{\mathcal{I}_S(u, v)\mathcal{I}^*{}_R(u, v)}{||\mathcal{I}_S(u, v)\mathcal{I}^*{}_R(u, v)||_2} \\ &= e^{-2\pi j(ut_x+vt_y)} \end{aligned} \tag{5.41}$$

The phase of the cross power spectrum can then be represented in the spatial domain by taking the inverse 2D Fourier transform of $\mathcal{PC}(u, v)$ i.e.:

$$PC(x, y) = \sigma(x - t_x, y - t_y) \tag{5.42}$$

where $PC(x, y)$ is an impulse function, located at (t_x, t_y) and the surrounding values will generally be order of magnitudes smaller, if not almost zero, as illustrated in Figure 5.9.

The phase correlation achieves a significant reduction in computational complexity especially when images are large and shows robustness against correlated and frequency dependent noise and changes in intensity due to varying illumination conditions [34]. Changes in illumination are usually slowly varying and therefore they are concentrated at low frequencies. PC is more robust against images that suffer from noise in narrow frequency bands, but it is not suitable for images which possess white noise across all frequencies or which have very low signal-to-noise ratio (SNR) at some frequencies [251, 300]. As opposed to CC, a distinct property of the PC method is the registration accuracy due to the detection of the peaks of the correlation function. The PC method tends to provide a sharp, distinct peak at the point of registration whereas classical CC methods can yield several broad peaks as well as a main peak whose maximum may

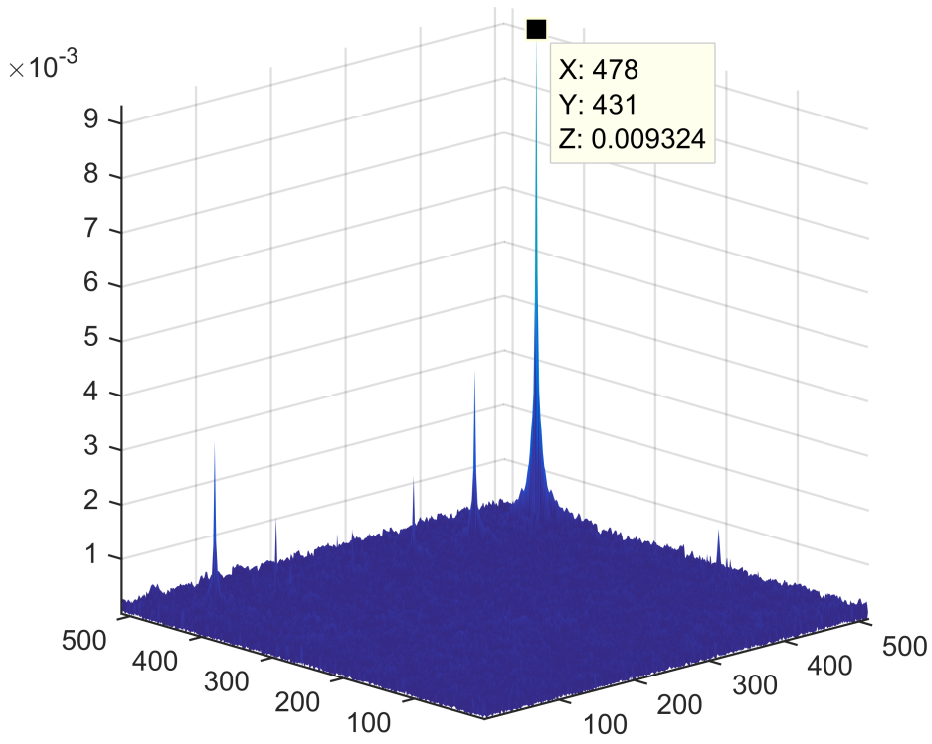


Figure 5.9: Illustration of phase correlation results using the images from Figure 5.7 showing the peak, and translation offset at [35,82].

not be exactly centred [92]. The grey-level CC peak can be broad but it is possible to interpolate to obtain sub-pixel resolution, which is not possible with PC methods.

In general, it can be seen that with area based registration techniques, the general four step framework described in section 5.1 were present, but were ambiguous and indistinct but in the feature based methods outlined in the next section the framework will be followed more concisely.

5.5 Feature Based Methods

Area based registration techniques operate by minimising dissimilarities between pixel intensity values between two images, a reference image I_R and a sensed image I_S. These techniques are often constrained to applications with small translations, rotations and scaling. The images must also have similar intensity functions [251]. Area based methods are preferably applied when images have little prominent details and the distinctive information is provided by greyscale/colour rather than by local shapes and structural information.

Feature based methods operate by extracting distinctive local spatial features such as points, edges, contours, lines and regions from both images [283], to match individual features to establish a global correspondence and to estimate the geometric transfor-

mation between the images [105, 250]. The extracted features consist of a control point or feature point, and a circular template [26]. The features should be distinct, efficiently detectable and expected to be stable in both images. Feature based registration methods are more robust to scene motion, changes in illumination and different sensor sources, which make them more suitable for a broader range of applications [34].

5.5.1 Feature Detection

The feature detection or feature extraction stage searches both the reference and sensed images for distinct features, that will provide robust matches, in order to be able to provide an accurate estimation of the transformation between the images. Therefore the detection of features is a crucial stage in the process. The features must aim to be distinct, robust to deformation/noise and efficiently detectable between images. The key properties for feature detectors are dependent on the application, but the most important properties are outlined below [117, 260]:

Robustness. The detection algorithm should be able to detect the same feature locations independent of scaling, rotation, shifting, photometric deformations, noise and compression artefacts.

Repeatability. The feature detection algorithm should be able to same features of the same scene or object repeatedly under varied viewing conditions

Accuracy i.e. should be able to accurately localise the image features (same pixel locations)

Generality i.e. should be able to detect features that can be used in different applications.

Efficiency. The feature detection algorithm should be able to detect features in new images quickly to support real-time applications

Quantity. The feature detection algorithm should be able to detect all or most of the features in the image. Where, the density of detected features should reect the information content of the image for providing a compact image representation.

Feature detection is a very research intensive field due to its broad applications in computer vision and image processing. In general, the types of features that can be detected within an image can be categorised as: point-like, line-like and region-like [18, 48, 230].

Point-like features can be described as pixels in an image which has strong variations in all directions [150]. Point-like features are rotation and translation invariant features [50] These point-like features can be represented as points themselves, end points,

corners, centres of line features or centres of gravity of regions [283]. Most techniques employed for detecting point-like features are based on the concept of the definition of a corner, which is usually defined as points of high curvature on the region boundaries [223, 293]. For example, Moravec's [193] interest point detector calculates a variance measure around a pixel within a window, the pixel is considered a feature if the variance measure exceeds a specified local maximum. The sum of squares of the differences of pixels, in each of the four directions i.e. horizontal, vertical, major and minor diagonals are calculated and the minimum value of the four is taken to be the variance measure. There are some limitations such as the method can only deal with a minimum rotational shift of 45°.

Harris [115] improved upon Moravec's corner detection by proposing the use of analytical expansions to deal with small rotational shifts, and uses a circular Gaussian window, rather than a rectangular window. The variance measure is defined as the principle curvatures of the second moment matrix of the Gaussian weighted patch of the image surrounding the pixel.

Beaudet [22] proposed a method to measure the cornerness at a pixel in an image by calculating the determinant of the Hessian Matrix, also known as the Hessian Detector. The work introduced in [63] used the Hessian detector to find the optimal corner model with two or three edges by an iterative gradient descent algorithm.

SUSAN (smallest univalue segment assimilating nucleus) [245] is a corner feature detector that uses a morphological approach rather than a differential approach which tend to be noise sensitive and computationally expensive. SUSAN classifies each pixel as a corner based on the intensity values within its surrounding neighbourhood. An efficient implementation of the SUSAN detector, known as FAST (features from accelerated segment test) considers a circular neighbourhood of a pixel and uses ID3 machine learning algorithm, which is a decision tree base for a pixel to be classified as a corner [150].

Line-like features are representations of general line segments [124, 273] such as object contours, man-made road or buildings etc. [17]. Line-like features are are most suitable for detecting man-made structures and of particular application to GIS (mapping) and satellite image registration. They are often expressed by the middle or end points of lines to establish correspondences [75, 267]. Standard approaches include using edge detectors such as first-order Canny edge detector [42], second-order based Laplacian of Gaussian (LoG) [172] and even curvature-based methods [62, 70, 106].

Lastly, region-like features sometimes called "blob" features, are closed boundary regions of an appropriate size, which tend to differ in properties such as brightness or colour in comparison to its surrounding regions [158]. These features are particularly suitable for object tracking applications [184] and for aerial imagery, especially in urban areas. The corresponding centres of gravity (i.e. the average location of an object

[187]) are control points used to represent region-like features as they are recognisable under rotation, scaling, skewing, random noise and grey-level intensity variation [300]. Some examples of region-like feature detection methods include Hessian-Laplace [187], Hessian-affine [161], Harris-affine [159], watershed transformation [157] and maximally stable extremal regions (MSER) methods [176].

The study of feature detection, extraction and descriptor assignment (methods to describe the details of a detected feature [35]) continues to be a very active area of research [250]. Point-like features often refereed to as interest points or control points have shown to match features with higher accuracy, as they are more invariant to changes in scale and affine transformations compared to line and region-like features [34, 106].

Scale Invariant Feature Transform

The scale invariant feature transform (SIFT) [165, 166] is a computer vision algorithm for detecting and extracting local features in an image. The method detects distinctive invariant features to perform reliable matching between different views of an object or scene based on high level considerations of the images structure. The detected features are chosen as such, so they are invariant to image scale and rotation, and robust to affine distortions, changes in illumination, changes in 3D viewpoint and noise. SIFT is broken down into four key stages:

1) **Scale-space extrema detection** In the first stage, a difference of Gaussian function (DoG) is used to efficiently search over all scales and image locations to identify potential interest points that are invariant to scale and rotation. This is accomplished using a continuous function of scale known as scale space [289]. Scale-space theory is a framework that represents an image as a family of blurred images with a variable scale, so that the scales can be detected [288]. It has been shown that under a variety of reasonable assumptions that the Gaussian function is the only possible scale-space kernel applicable [144,158], therefore for an image, $I(x, y)$ the scale space of an image, $L(x, y, \sigma)$, where σ is the representative image scale, can be defined as:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{5.43}$$

where $*$ is the convolution operator and $G(x, y, \sigma)$ is the variable-scale Gaussian kernel, defined as [160, 166, 171]:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{5.44}$$

It has been demonstrated by detailed experimental comparisons [186], that the maxima and minima of the scale normalised Laplacian of Gaussian, $\sigma^2\nabla^2 G$ produce the most stable image features compared to a range of other approaches (e.g. Hessian-affine [161], Harris-affine [159] or Harris-Laplace [185]. Additionally, it has been shown that the DoG function is approximately equal to the scale normalised Laplacian of Gaussian $\sigma^2\nabla^2 G$ [159]. Thus, to efficiently detect the location of stable interest points, the extrema of the DoG function can be found from the difference between two consecutive scale-space representations of scales σ and $k\sigma$, where k is a multiplicative factor representing the difference of scales [166]:

$$D_g(x, y, \sigma) = L_s(x, y, k\sigma) - L_s(x, y, \sigma) \tag{5.45}$$

An efficient computational approach to detect the local extremum, $D_g(x, y, \sigma)$ has been proposed in [185], using an image pyramid approach as illustrated in Figure 5.10. Initially, several octaves of scale-space images are generated from the original image by incrementally convolving the initial image with Gaussian, producing images separated by a constant factor k in scale space (i.e. the left hand column in Figure 5.10). Each octave is half the image size of the previous octave (i.e. doubling of σ). In order to divide each octave into an integer number s of intervals, k should be chosen so that $k = 2^{1/s}$. Furthermore, in order to search a complete octave for final extrema detection, $s + 3$ images in the stack of blurred images for each octave must be produced. Adjacent image scales are then subtracted to produce the difference of Gaussian images, shown in the right hand column in Figure 5.10.

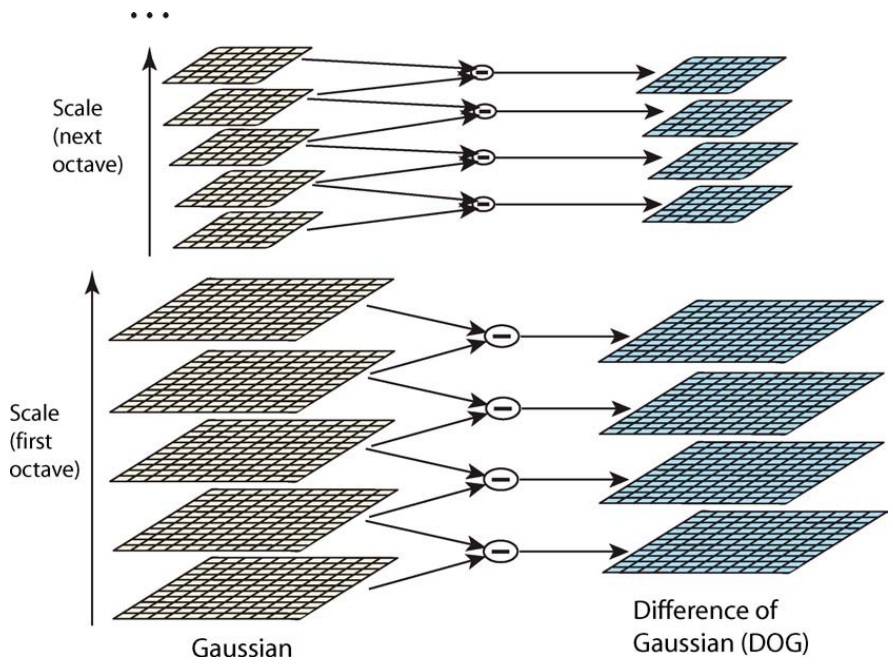


Figure 5.10: For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference of Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated [166].

2) **Keypoint localisation** For each interest point candidate, a detailed fit for location, scale and ratio of principal curvatures with respect to the nearby data is modelled. This information allows points to be rejected which have low contrast, making them susceptible to noise or are poorly localised along an edge. It is possible for an extrema to be located at a sub-pixel location, it is reasonable to locate them precisely. This can be achieved by employing Taylor expansion (up to the fourth quadratic term) to the DoG image [36]:

$$D_g(\mathbf{x}) = D_g + \frac{\partial D_g^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D_g}{\partial \mathbf{x}^2}\mathbf{x} \tag{5.46}$$

where D_g and its derivatives are evaluated at the candidate interest point and $\mathbf{x} = (x, y, \sigma)^T$ is the offset from the interest point. The location of the extremum, $\hat{\mathbf{x}}$, can be determined by taking the derivative of the function above (eqn. 5.46) with respect to $\hat{\mathbf{x}}$ and setting it to zero, giving:

$$\hat{\mathbf{x}} = -\frac{\partial^2 D_g^-1}{\partial \mathbf{x}^2}\frac{\partial D_g}{\partial \mathbf{x}} \tag{5.47}$$

If the resulting offset $\hat{\mathbf{x}}$ is greater than 0.5 in any dimension, then it means that the extremum lies closer to a different interest point. In order to obtain the

final $\hat{\mathbf{x}}$ offset, the interest point is changed and the process above is repeated, until the final offset value is obtained. Once the final offset value is found, the corresponding function value $D_g(\hat{\mathbf{x}})$ is calculated. If the function value is below a specified threshold value, then the interest point is rejected due to its low contrast properties. This function value can be found by substituting eqn. 5.46 into eqn. 5.45, resulting in:

$$D_g(\hat{\mathbf{x}}) = D_g + \frac{1}{2}\frac{\partial D_g^T}{\partial \mathbf{x}}\hat{\mathbf{x}} \tag{5.48}$$

For stability it is not sufficient to reject interest points based on low contrast alone. The DoG function will have strong responses along edges, even if the location of the edge is poorly determined [166]. A poorly defined peak in the DoG will have a large principal curvature across an edge but a small principal curvature in the perpendicular direction. The principal curvatures can be computed from a 2×2 Hessian matrix, $\mathbf{H}$ calculated at that location and scale of the interest point.

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{5.49}$$

the derivative can be determined by taking the difference of the neighbouring sample points. The eigenvalues of $\mathbf{H}$ are proportional to the principal curvature of D_g. From [115], it is possible to avoid explicitly calculating the eigenvalues, as it is the ratio of the principal curvatures that is more important. If the ratio of the eigenvalues are below a set threshold the interest point is rejected [166].

3) **Orientation assignment** At this stage, the aim is to assign an orientation to the chosen interest points based on the local image gradient direction. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale and location for each feature providing invariance to the transformations [166]. By assigning a consistent orientation to an interest point based on the local image gradient, invariance to rotation is achieved. [166] present an approach which through experimentation found to provide the most stable results. For each image sample $L(x, y)$, within a region around the interest point, located at (x, y), the magnitude of the gradient, m is calculated:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \tag{5.50}$$

The orientation of the interest point and its surrounding region can be computed as:

$$\theta(x, y) = \tan^{-1} \frac{(L(x, y+1) - L(x, y-1))}{(L(x+1, y) - L(x-1, y))} \tag{5.51}$$

From the computed gradient orientation an orientation histogram is formed with 36 bins covering a 360° range of orientations. Each sample is weighted dependent on its gradient magnitude, m and by a Gaussian weighted circular window with a σ that is 1.5 times of the scale of the interest point. From the histogram, the global peak is located and assigned to the interest point along with any other peak orientations that are within 80% of the height of the global peak. For, locations with multiple peaks of similar magnitude, multiple interest points will be created at the same location and scale but with different orientations. Multiple orientation interest points have been shown to significantly contribute to the stability of the matching [36]. A parabola is then fitted to the three histogram values closest to each peak to interpolate the peak position with better accuracy [165].

4) **Keypoint descriptor** Often referred to as interest point descriptor or feature descriptor is a unique descriptor is assigned to each keypoint, so they can be distinguished from each other. A window of 16×16 centred around each keypoint is taken and then further broken down into 16 windows of size 4×4. Within each 4×4 window, the gradient information (i.e. the gradient magnitude and gradient orientation) of all pixel locations are calculated. These orientations are placed into an 8 bin histogram resulting in a 128 element, feature vector, as shown in Figure 5.11

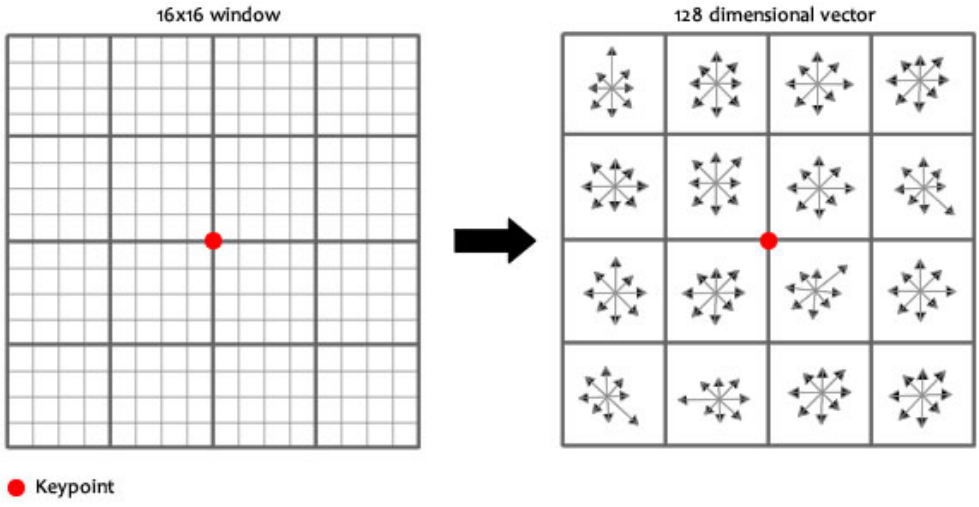


Figure 5.11: A 16×16 SIFT window, centred around each keypoint, then further broken down into 16 windows of size 4×4. Within each 4×4 window, the gradient information (i.e. the gradient magnitude and gradient orientation) of all pixel locations are calculated. These orientations are placed into an 8 bin histogram resulting in a 128 element [243]

Speeded Up Robust Features

Speeded up robust features (SURF) [20, 21] is a local feature detector and descriptor algorithm influenced by SIFT. While in terms of performance, both techniques are similar, computationally SURF is considerably faster [204].

In SURF, the Laplacian of Gaussian (LoG) is approximated using a box filter rather than the difference of Gaussian (DoG) as employed by SIFT. This modification offers two advantages: First, the convolution with a box filter can be obtained easily with the utilisation of integral images. Secondly, the processing can be done simultaneously for different scales. SURF relies on the determinant of the Hessian matrix to determine both scale and location, which for many applications is sufficient, but additionally SURF offers the option to compute the orientation, resulting in computational efficiency, by the removal of unnecessary calculations [252].

For orientation assignment (if required), SURF uses the Haar wavelet responses in the horizontal and vertical directions for a circular neighbourhood of radius $6s$ around the interest point with s the scale at which the interest point was detected. Once the wavelet responses have been calculated the responses are weighted with a Gaussian ($\sigma = 2s$) centred at the interest point [20]. The dominant orientation is estimated by the sum of all the responses within a sliding orientation window of $60°$. The Haar wavelet responses can be obtained easily with the use of integral images. For applications in which rotation invariance is not required, this stage is skipped, this version of SURF is often called upright-SURF or U-SURF, which is computationally efficient and robust up to $\pm 15°$ [252].

For interest point description, SURF uses wavelet responses in the horizontal and vertical directions (once again). First a region of size $20s \times 20s$ is centred around the interest point where s is the scale at which the interest point was detected. This is then divided into 4×4 subregions. Each subregion, horizontal (d_x) and vertical (d_y) Haar wavelet responses are taken, and a vector (ν), is formed to describe the interest point:

$$\nu = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right) \tag{5.52}$$

The result is a feature (interest point) descriptor with 64 dimensions. This can be extended to 128 dimensions for more distinctiveness by computing the sums of the elements (i.e. $d_x, d_y, |d_x|, |d_y|$) separately for positive and negative wavelet responses in the horizontal and vertical directions. SURF can also be extended to operate on colour images (C-SURF) providing further distinctive feature descriptors and increasing the dimensionality [97]. It is important to note that as the dimensionality increases the matching reliability improves but at the cost computational time. SURF uses the sign of the Laplacian (trace of the Hessian matrix) to indicate whether an interest point is detected from a bright blob on a dark background or conversely i.e. dark blob on bright background and with no additional computation cost as it is already computed

during detection. In the matching stage the feature descriptors detected from the same situation are compared only, which results in faster matching, without reducing the descriptors performance [139, 150, 252].

5.5.2 Feature Matching

The next stage in the image registration process is to match the detected salient features. As discussed earlier, area based methods deal with matching using a predefined sized window or even the entire image. In the case of feature based methods the interest points are estimated for a perfect match between the reference and sensed images. Feature based methods focus on the spatial relations or various descriptors of features [26]. The simplest approach would be to match each individual feature from the sensed image to the reference image, using some form of distance calculation (i.e. Euclidean, Hamming [108], Mahalanobis [186]), by minimising some cost function [213] or another form of similarity measure (i.e cross correlation) [34]. There are many sophisticated matching techniques such as those proposed in [16, 28, 200, 214, 256, 257]. For example, in [16] Barrow introduced the chamfer matching technique for image matching. In this technique line features detected in the images are matched by the minimisation of the distance between the features. Recently, Wen [283] developed a high performance feature matching technique for image registration by combining spatial and similarity information. It is important to note that many extracted features may not have a corresponding match, due to different sensors, different acquisition conditions, or simply that similar features were not detected in both images. The more sophisticated matching techniques tend to be computationally expensive and unsuitable for some applications (i.e. video stabilisation, object tracking) therefore more efficient algorithms focus on the concept of finding the nearest neighbours in the feature space. Two widely used solutions to the feature matching problem are RANSAC [85] and least median of squares (LMS) [226].

RANSAC

Random sample consensus (RANSAC) [85] is a general parameter estimation approach algorithm designed to mathematically model input data that contains outliers [64]. RANSAC functions by assuming all data consists of either inliers that can be described by a mathematical model with a set of parameters within some error threshold which defines the maximum deviation attributable to the effects of noise. The percentage of outliers that RANSAC can handle, also known as the breakdown point can be greater than 50% but is often assumed to be the practical limit for many other commonly used techniques such as the least median of squares [301]. There exists a robust estimator proposed by [248], which is capable of estimating the parameters of a model using datasets containing more than 50% outliers. Although RANSAC is able to estimate

the model parameters with high accuracy even when there exists a significant number of outliers, it may not be able to obtain the optimal parameters when the number computational iterations is insufficient [85].

In general, there are several slight variants [51, 52, 175, 195, 233, 259, 297] to the original RANSAC algorithm [106], but the essential algorithm is composed of a two step iterative framework (hypothesis-and-test framework):

1. Hypothesise - A minimal subset from the input data is randomly selected. The random selection of data offers two advantages: there is no need to employ complex optimisation algorithms. Furthermore there is no need for large amounts memory [49]

2. Test - In this step the algorithm checks which elements of the entire database are consistent with the predicted model parameters obtained in the previous step. If a data element does not fit the predicted model within some error threshold, it is considered an outlier.

The set of inliers obtained for the fitting model, in the first step is called consensus set. The above two step procedure is iteratively repeated until the obtained consensus set in a certain iteration has enough inliers to provide a confident estimation of the model parameters.

RANSAC is initialised by randomly selecting a subset of k feature correspondences, which is used to estimate the transformation model parameters. The transformation is then applied to the full set of feature correspondences to obtain the errors i.e. the differences between the mapped estimated locations and the detected feature locations. RANSAC then determines the number of inliers exist based on a predefined error threshold (typically around $1-3$ pixels [250]). The above process is repeated S_{ransac} times and the estimated model parameters from the feature correspondence set with the greatest number of inliers, is taken as the best solution.

An important factor to consider is the number of iterations, S_{ransac}, that must be processed. S_{ransac} must be sufficiently large so that the probability P_{ransac} of finding an accurate corresponding set of inliers is highly likely (typically $P_{ransac} \approx 0.99$ [249]). Let p_{ransac} represent the probability that any randomly selected feature correspondence is an inlier. Then, the likelihood for an iteration where all k randomly selected feature correspondences are inliers is p_{ransac}^k. The likelihood that all S_{ransac} iterations fail is [249]:

$$1 - P_{ransac} = (1 - p_{ransac}^k)^{S_{ransac}} \tag{5.53}$$

with some mathematical manipulation, the minimum number of iterations required can be obtained to be [249]:

$$S_{ransac} = \frac{\log(1 - P_{ransac})}{\log(1 - p_{ransac}^k)} \tag{5.54}$$

Least Median of Squares

Least median of squares (LMS) [226] is another widely used robust model estimation technique used for feature matching. It operates similarly to RANSAC but instead of counting the number of outliers on the basis of whether the corresponding errors are within a predefined error threshold, the median value of these errors are calculated and then an error threshold for the median value is minimised. LMS fits outliers for robust regression as a minimisation problem with a non-linear loss function, thus making it necessary to employ a numerical optimisation technique. Unlike RANSAC, LMS performs poorly under circumstances where more than 50% of the features are outliers, but has the advantage that no error threshold value is required.

5.5.3 Transformation Model Estimation

Once the outliers have been detected and removed, the next stage involves estimating the transform between the matching feature correspondences. Accurately estimating the parameters of the transformation model between the sensed and reference image minimises the registration error once the sensed image has been transformed. In general, the number of matched feature correspondences available is greater than the minimum needed to estimate the transformation model, for example a projective deformation would require a minimum of 4 matched feature correspondences, hence this is an overdetermined case where there are more equations than there are unknowns.

Least Squares

The classical and most widely used method is the least squares method. The least squares method approximates the solutions of an overdetermined system of equations. The least squares approach minimises the sum of squared error the matched feature correspondences. For the most general case i.e. projective transformation (Eqn. 5.30-5.31) can be rearranged as [151]:

$$-(h_{11}x + h_{12}y + h_{13}) + (h_{31}x + h_{32}y + 1)x' = 0 \tag{5.55}$$

and

$$-(h_{21}x + h_{22}y + h_{23}) + (h_{31}x + h_{32}y + 1)y' = 0 \tag{5.56}$$

The coefficients of the homography matrix appear linearly and therefore the above equations can be rearranged [250, 262]:

$$\mathbf{a}_{h_x}^T \mathbf{h} = 0 \tag{5.57}$$

and

$$\mathbf{a}_{h_y}^T \mathbf{h} = 0 \tag{5.58}$$

where

$$\mathbf{h} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, 1]^T \tag{5.59}$$

$$\mathbf{a}_{h_x} = [-x, -y, -1, 0, 0, 0, xx', yx', x']^T \tag{5.60}$$

$$\mathbf{a}_{h_x} = [0, 0, 0, -x, -y, -1, xy', yy', y']^T \tag{5.61}$$

The following linear system of equations can be formed for a given set of matched feature correspondences:

$$\mathbf{A}_h \mathbf{h} = 0 \tag{5.62}$$

where

$$\mathbf{A}_h = \begin{bmatrix} \mathbf{a}_{h_x}^T \\ \mathbf{a}_{h_x}^T \\ \vdots \end{bmatrix} \tag{5.63}$$

A least squares estimate can be obtained by calculating the eigenvector corresponding to the minimum eigenvalue of $\mathbf{A}_h^T \mathbf{A}_h$.

Weighted Least Squares

The least squares method assumes that all the matched feature correspondences, are matched with the same accuracy, hence they are averaged equally over the entire image. This means it cannot account for local geometric distortions such as sensor non-linearities, atmospheric conditions and local 3D scene features observed from different viewpoints [34]. To negate this problem a weighted least squares approach is often employed [164]. Weighted least square methods generally associate the weights with an estimation of variance of the local region surrounding the matched feature correspondences [166].

5.6 Image Transformation/Resampling

Image resampling is the process of geometrically transforming an image [66]. Once the transformation model has estimated the mapping function (previous step) required to register the sensed image to the reference image, the task of applying the mapping function to the sensed image must be considered. In many image processing applications not limited to image registration, images can be transformed, so that a pixel in the sensed image may be mapped to a fractional pixel location. To alleviate the effect of fractional pixel locations, several interpolation techniques have been proposed. An interpolation technique estimates the values at unknown points using the intensity values from known neighbouring pixel locations [106]. Different methods have been developed to estimate the pixel intensities at unknown locations. In the following sections three of the most widely used methods will be discussed: nearest-neighbour, bilinear and bicubic interpolation.

Nearest-Neighbour

The nearest-neighbour method is the most simplest of interpolation algorithms. This method only considers a single neighbouring pixel (i.e. the nearest known pixel location) and assigns that intensity value to the unknown pixel in question. In terms of performance, the nearest-neighbour algorithm is very simple and therefore outperforms other interpolation techniques in terms of processing time. This method generally suffers from artefacts such as blurring and jaggedness due to it simplistic interpolation [217, 253].

Bilinear Interpolation

Bilinear interpolation computes the new pixel intensity values using linear interpolation. The method operates by considering the 2×2 neighbouring pixel values surrounding the unknown pixel. It then calculates the weighted average of these 4 pixels to obtain the final intensity value. This results in a much smoother image compared to the nearest-neighbour [229].

Bicubic Interpolation

Bicubic interpolation computes the unknown pixel intensity values using a cubic function to estimate the intensity value rather than a linear function, as compared to bilinear interpolation. In order to provide a good estimate the bicubic method requires a slightly larger 4 region of known neighbouring pixel intensity values. The method considers the surrounding 4 pixels and then generates a weighted average of the 16 pixels to obtain an estimate for the unknown interpolated pixel intensity value. This method produces sharper, smoother images compared to the previous methods described above, with reduced blurring and "blocky" artefacts [229].

5.6.1 Implementation

To enable image registration between frames, for our scenario, it was decided that the speeded-up robust features (SURF) algorithm will be used to detect blob features and then RANSAC will be used to match the corresponding features across adjoining images and the reference image. SURF and RANSAC both provide accurate feature correspondence as well as feature matching and transform estimation, with the least computational time and good registration performance. Additionally, once the image registration was implemented, a brief study on the potential bandwidth savings was conducted and reviewed in Appendix A.

5.7 Chapter Summary

A comprehensive overview of image registration techniques have been outlined, ranging from global area based methods to localised feature based methods. Area based methods are very robust to affine transformations, but struggle when dealing with large rotations or the presence of perspective transforms. Area based methods can be computationally expensive to implement, especially on high resolution images, as they tend to be global methods. Methods to make them computationally cheaper such as localised area-based methods have been developed, such as template matching, but they can still be inefficient with large rotations, and are still ineffective against perspective deformations. Feature based methods are localised methods which operate on a large number of local regions of an image, to find matching features sets between feature correspondences and using statical analyses to determine the transform between the images. Feature based methods operate computationally more efficiently than area based methods, and can function with perspective transforms. For the application to the novel imaging system, the computational efficiency of feature based methods, along with ability to operate under perspective deformations provides the most suitable choice for our application to stream a live video feed, from an aircraft system, to a remote terminal operator using the proposed dual-field-of-view imaging system. SURF has been shown to operate with similar registration performance to SIFT, but with better computational efficiency, with the use of the RANSAC algorithm to estimate the transform between frames. The implementation of image registration enables the work outlined in Chapter 4 to be extended to operate over sequential frames.

Chapter 6

Temporal Averaging Geometric Separation Technique

6.1 Introduction

Missile systems and uninhabited (or unmanned) aerial vehicles (UAVs) are being used to provide surveillance imagery and reconnaissance information in a number of environments where human operators cannot operate safely. Although, both missile systems and UAVs were developed initially for military applications, there are a number of other situations where the remote collection of aerial imagery would be very useful; e.g. for civilian fire and rescue services, and the inspection of nuclear and chemical facilities. However, a lot of these aircraft have limited payload capability, power budget and endurance. High quality imaging systems are often heavy, and they require significant processing, electronic storage and communications equipment to capture and transmit imagery to a remote operator or ground control station. This chapter considers a technique that could reduce the requirement for an aircraft to have multiple imaging systems providing different capabilities; providing both situational awareness and image stabilisation information from a wide field-of-view camera, and detailed object identification information from a narrow field-of-view camera. The proposed sensor would image both wide field-of-view ($wFoV$) and narrow field-of-view ($nFoV$) at the same time, and focus the images onto the same focal plane array, thereby forming a single image with both scenes superimposed on the other (i.e. an equal mixture of both images). The proposed technique would then allow the two scenes to be recovered by post-processing the superimposed image using the geometric relationship between the wide field-of-view and the narrow field-of-view. Initial results for this geometric technique were reported in [179]. In this chapter, the resultant images are filtered in the time-domain to improve the quality of the image separation, and the results are compared against a standard benchmark approach based on image super-resolution using conventional images. It is demonstrated that the performance of the time-averaged geometric separation method is similar to that of the ideal super-resolution process

but is significantly more robust to errors in registration of sequential images. The time-domain filtering process extends the geometric separation technique to sequential frames, i.e. live streaming videos. This will be of particular interest for applications involving an ongoing surveillance of an area of interest. The situational information obtained from the wide field-of-view allows for continuous monitoring of the general area, whilst the narrow field-of-view allows for closer interrogation and identification of objects of interest.

Imaging systems have many uses in aerial vehicles; providing imagery for pilot flying aids, airborne surveillance and reconnaissance, target search, detection, identification and localisation, and visual navigation or photogrammetry. The different requirements for airborne imaging are often contradictory. Pilot flying aids, navigation and reconnaissance require wide-angle images and the ability to cover large areas of the ground. Target search, detection and identification require reduced fields of view, obtaining more detailed imagery of particular areas of the ground or specific objects. To provide all of these functions, conventional military aircraft often have several electro-optical imagers, which work in a coordinated way so that a wide field-of-view camera cues a narrow field-of-view imager to interrogate possible threats or targets.

As has been noted, missile systems, as well as UAVs (particularly lightweight UAVs) are often limited in terms of payload, power and endurance; so having multiple camera systems is often not a practical solution to address this problem. Other approaches that have been considered to provide both wide field-of-view situational awareness and narrow field-of-view high resolution imagery for object detection and identification have tended to use variable resolution imagers similar to the human vision system (foveal imagers) [12, 292], or post-processing the images to extract sub-pixel information from a sequence of imagers (super-resolution) [39, 40, 101, 119, 271, 274]. Foveal imagers have a small area at the centre of the field-of-view with a high density of pixels and a lower density of pixels outside this region [12, 292]. Super-resolution techniques rely on the alignment of a set of images of the same area to sub-pixel accuracy and then estimating the underlying fine structure using the slight variations in the pixel intensity values within each overlapping pixel [29, 83, 102, 145, 205].

The aim of this chapter is to discuss an alternative approach, where a single sensor system is used to image two separate images simultaneously, one wide field-of-view and one narrow field-of-view, and the resultant superimposed image is separated by post-processing that utilises a geometric relationship between the two individual images. The result is a wide field-of-view image that can provide contextual information for surveillance and/or image stabilisation, and a narrow field-of-view for object identification and localisation. The use of superimposed images has the additional benefit of potentially reducing the need to transmit multiple image streams from a UAV to a ground control station. The base images used in this work are examples only, and not

sequences derived from aerial surveillance and reconnaissance because of the difficulty in obtaining suitable open source surveillance and reconnaissance images. One of the images is an open source aerial image (looking down at the ground), Image [A], and one is an image of a natural scene taken with a modern digital camera (looking horizontally at vertical structures), Image [B].

6.2 Temporal Geometric Separation

The separation of multiple superimposed images has been considered in a number of areas (e.g. reflections and semi-transparent surfaces [33, 82]) and have used techniques adapted from signal processing utilising the higher-order statistics of the images and independent component analysis [43, 55–57]. This section describes a technique that is an extension of a technique proposed in [179], where two superimposed images were separated using a known geometric relationship between the two images. It describes an extension that utilises a sequence of images (similar, in essence, to super-resolution methods) to reduce the artefacts and noise introduced by the single-frame geometric separation technique.

The technique presented in Chapter 4 is known to suffer from image noise effects or macro-blocking issues, as discussed in [179]. To address this problem, a time-averaging technique is implemented to filter out the artefacts created by the geometric separation on single images. This has been simulated by applying a random sub-pixel shift to the captured image of the scene, prior to the images being superimposed. This is equivalent to taking a sequence of images with small displacements of the camera from frame to frame. The images can either be separated on-board, or (preferably) after transmission to a remote operator/ground control station. The single frame separation is applied to each superimposed image, and the two components aligned/registered to a common reference image and then time-averaged.

6.3 Super-resolution

Super-resolution (SR) image reconstruction produces a high-resolution image from a set of low-resolution images [29,39,40,83,101,102,119,145,205,271,274]. There are two main approaches to super-resolution; the first is a frequency domain approach, which tends to be computationally cheap but highly sensitive to translational motion [29]. This sensitivity can degrade performance substantially. The alternative approach is a spatial domain approach, which is more robust but can be computationally expensive. In this case, the lower resolution images are aligned using image registration techniques before the images are combined to form a high resolution super-resolution image. In both cases, small (sub-pixel) motion from frame-to-frame provides additional spatial information to allow fine spatial structure within the scene to be reconstructed even

though the resolution of each individual image is limited by the size of the pixels.

A spatial super-resolution approach is used as a benchmark for the proposed time-averaged geometric separation method. A sequence of low-resolution (spatially averaged) 512×512 images is created from the original high-resolution image (representing a conventional camera). Each image in the sequence has a small translational shift in horizontal and vertical directions before the image is scaled down. This provides the sub-pixel information used by the super-resolution method. The super-resolution reconstruction technique takes each 512×512 low-resolution image and scales it up to 2048×2048 pixels. The images are then aligned to correct for the translational motion and averaged to form a single high-resolution image.

Super-resolution is used as a benchmark technique because, it has been studied previously for UAV surveillance applications [39, 40, 101, 119, 271, 274] and it provides a best case for comparison. To ensure that the comparison is against an optimal benchmark, the image registration method used to correct for the translational shift is assumed to be perfect initially, and registration errors are added later to examine the robustness of each of the approaches.

6.4 Results

The images produced by each technique are analysed quantitatively using standard image quality metrics (IQM), averaged pixel intensity error and the structural similarity index (SSIM) [276]. The averaged pixel intensity error metric calculates the percentage difference in image contrast, between the original source image and the resultant image obtained from the respective techniques. The SSIM index assesses the image quality with respect to image contrast, luminance and structural information, providing a broader, more accurate representation of image quality [276].

6.4.1 Perfect Image Registration

The results obtained for comparison are:

{1} A benchmark super-resolved image, I_1

{2} A time-averaged $nFoV$ image, I_2

{3} A $nFoV$ image separated from the time-averaged $wFoV$ image, I_3

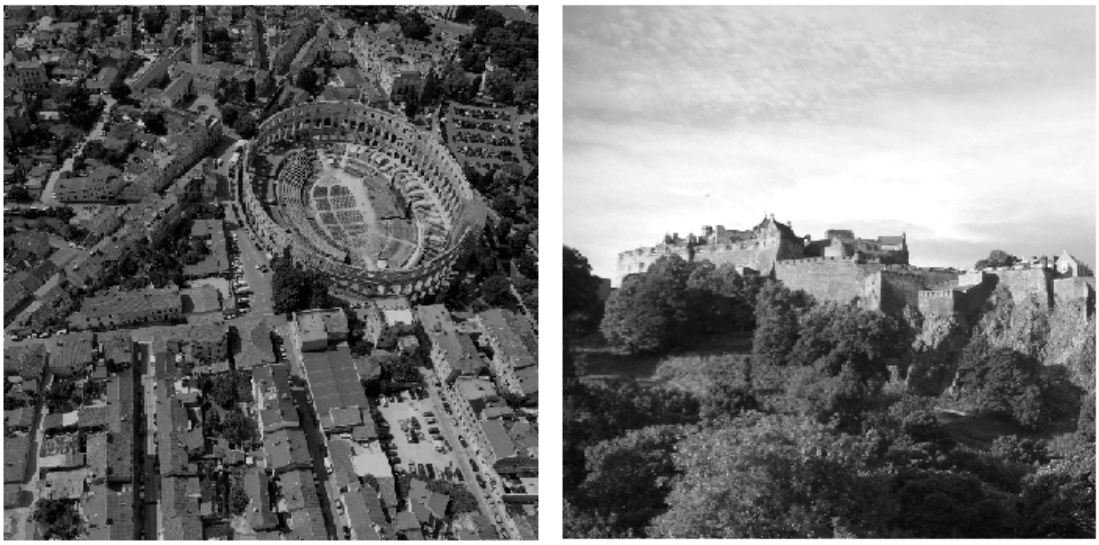
{4} A time-averaged $wFoV$ image, I_4

The results for {3} are obtained by filtering the separated $wFoV$ image {4}, and separating the $nFoV$ image by calculating:

$$I_3 = I_m - cI_4/d \tag{6.1}$$

where I_m is the superimposed and the constants, c and d are the weightings (0.5 in our case), as defined in Chapter 4.

These results have been compared against the original input images (Figures 6.1 and 6.3) by calculating the error in the image intensity, averaged over all pixels in the image. The structural similarity (SSIM) index has also been used since it more accurately reflects the structural similarity of the source and input images as well as considering the contrast and luminance properties when determining the image quality [276].

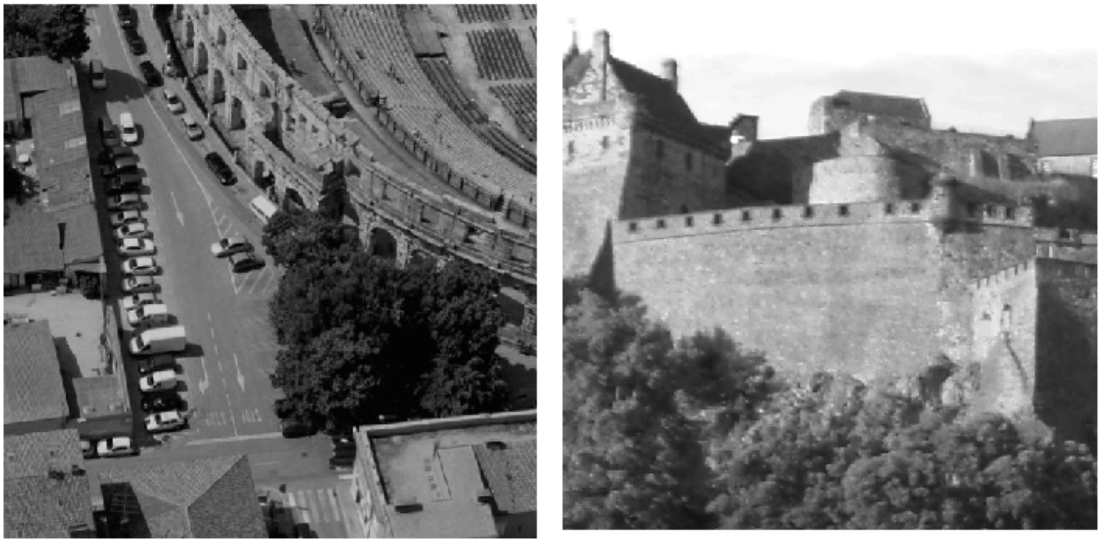


(a) Image [A] (b) Image [B]

Figure 6.1: 512×512 $wFoV$ sample images

(a) Image [A] (b) Image [B]

Figure 6.2: 512×512 superimposed images



(a) Image [A]

(b) Image [B]

Figure 6.3: 512×512 $nFoV$ images

In this study, the results were calculated over a sequence of 50 image frames, after which there was no significant improvement to the results. Intermediate results were recorded and averaged for different sequences and sub-pixel shifts. The averaged results are shown in Figures 6.4 - 6.5 for sample image [A] and Figures 6.6 - 6.7 for sample image [B]. Once the sequence of 512×512 images have been created, the original 2048×2048 high-resolution images are used for comparison purposes only.

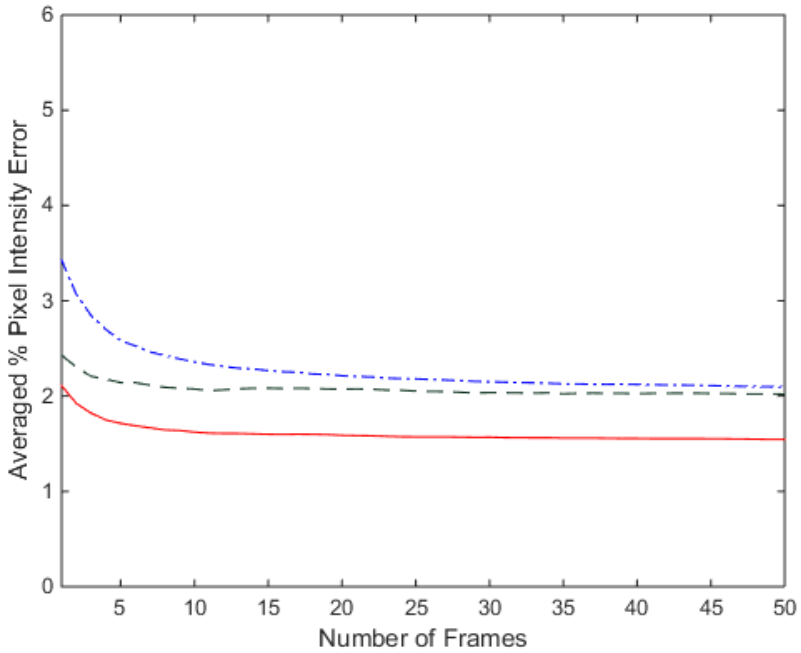


Figure 6.4: Averaged pixel intensity error values calculated for the $nFoV$ images 1-3 and the original $nFoV$ image (Fig 6.2a), using the digital camera image (Image [A]): 1 = red/solid, 2 = green/dash, 3 = blue/dot-dash.

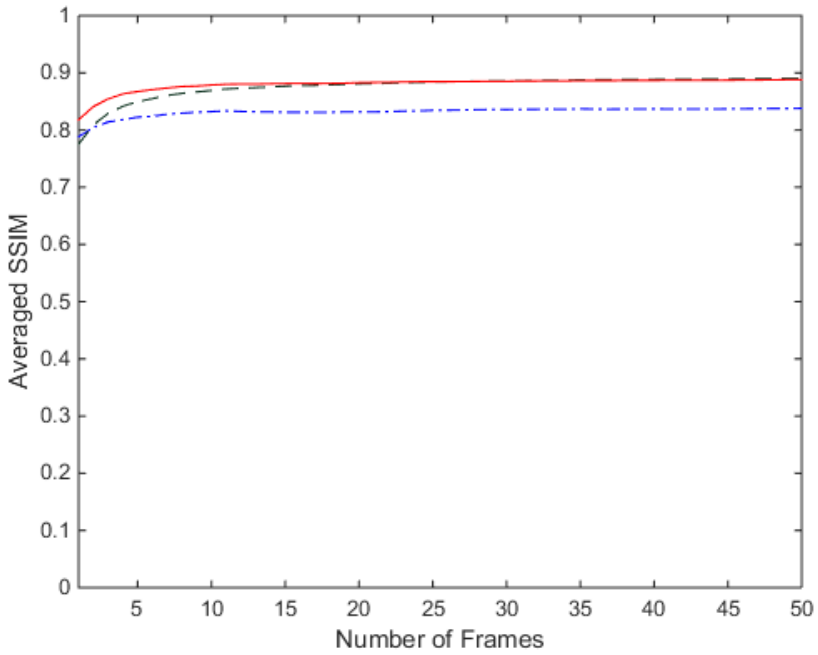


Figure 6.5: Averaged SSIM values calculated for the $nFoV$ images 1-3 and the original $nFoV$ image (Fig 6.2a), using the digital camera image (Image [A]): 1 = red/solid, 2 = green/dash, 3 = blue/dot-dash.

For the work presented, a scaling ratio, $\gamma = 4$ has been used for the time-averaged geometric separation technique. Super-resolution is applied to images that have random sub-pixel translations to each of the input images prior to reduction to 512×512 pixels. The same random shift is applied to the base image before the superimposed images are created for the geometric separation technique. The images are superimposed according to the process outlined in [179]. The small misalignments introduced between frames are removed using image registration to sub-pixel accuracy and aligned to a common reference image as would be the case for conventional super-resolution. Although for the proposed time-averaged separation technique, the high frequency structural information is already present within the $nFoV$ component of the superimposed image. The sequence of separated images are then filtered (averaged) in the time domain.

Figure 6.4 presents the average number of differences (pixel intensity errors, expressed as a percentage) between the output images {1}, {2} and {3} against the corresponding input image. Figure 6.5 gives the corresponding averaged SSIM values for the same images.

From Figure 6.4, it can be seen that the super-resolution technique {1} has the least significant average pixel intensity error, but the filtered $nFoV$ image {2} also contains a low value for the averaged pixel intensity error, with approximately 1% difference between the two methods. Once $5 - 10$ images are processed, it can be seen that there is very little improvement in the image quality. The performance of the $nFoV$ image {3} generated by separation from the filtered $wFoV$ image is noticeably lower than the

other two results.

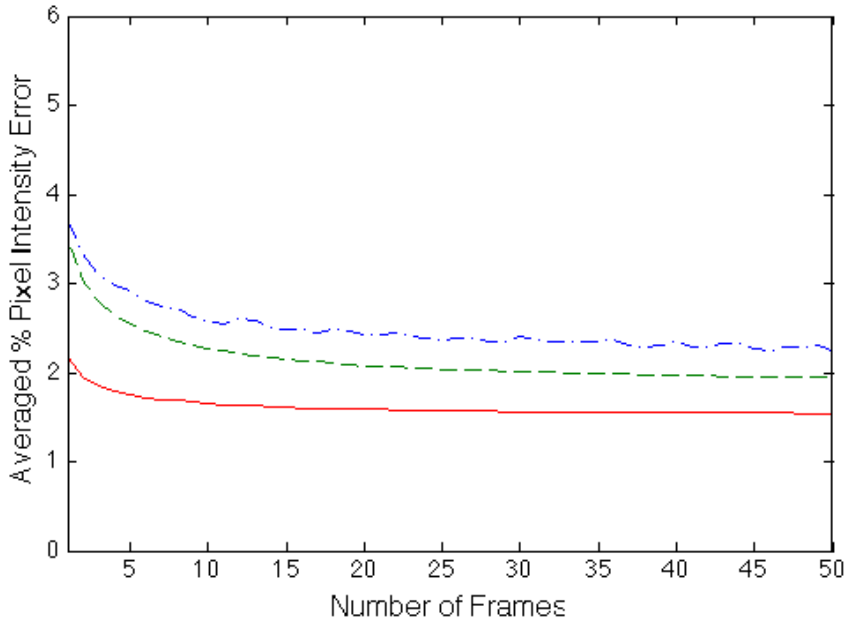


Figure 6.6: Averaged pixel intensity error values calculated for the $nFoV$ images $\{1\} - \{3\}$ and the original $nFoV$ image (Fig 6.2b), using the digital camera image (Image [B]): 1 = red/solid, 2 = green/dash, 3 = blue/dot-dash.

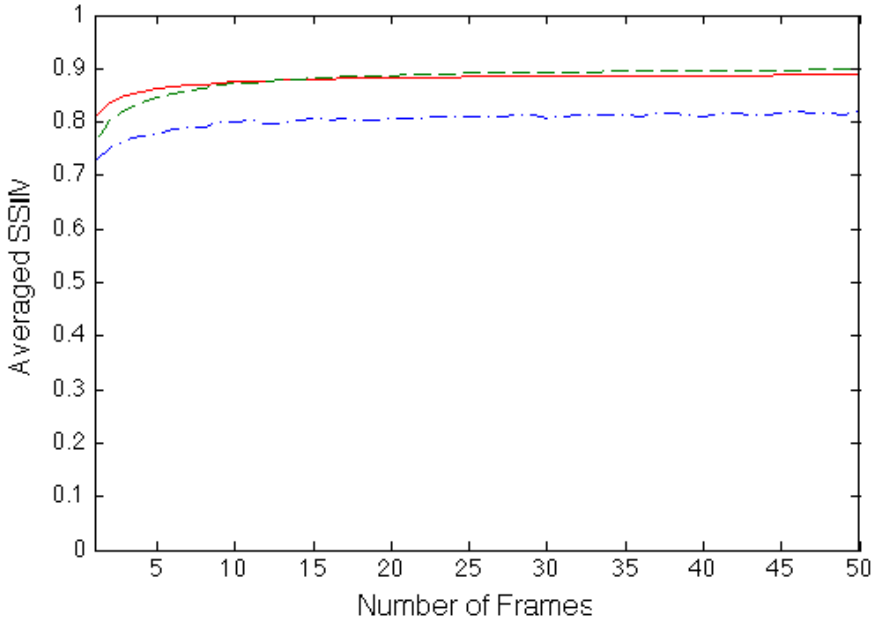


Figure 6.7: Averaged SSIM values calculated for the $nFoV$ images 1-3 and the original $nFoV$ image (Fig 6.2b), using the digital camera image (Image [B]): 1 = red/solid, 2 = green/dash, 3 = blue/dot-dash.

Figure 6.5 shows the structural similarity (SSIM) [276] of the results obtained from the geometric separation technique and the super-resolution technique compared

against the original input images (Figure 6.1 & Figure 6.3). The original super-resolution technique {1} gives the best results, although marginally, for a small number of images. The filtered $nFoV$ image {2} provides similar levels of performance after about $5-10$ images have been processed. The time-averaged $nFoV$ initially {2} gives a structural similarity of 75% but within 10 frames, this significantly increases to almost 90% similarity. The $nFoV$ generated from the filtered $wFoV$, {3}, shows the same trend as {1} and {2} but the structural similarity is much lower compared to the other results. It is worth highlighting that the filtered $nFoV$ geometrically separated image {2} provides very similar structural image content (as measured by the SSIM) even though the pixel intensity errors are higher on average. This could be due to the parameters used by each image metric to assess the image quality. The averaged pixel intensity error percentage only considers the pixel-by-pixel contrast error, whilst the SSIM not only considers contrast but also luminance and structural information.

Figure 6.8: Single separated $nFoV$ images, showing image noise arising from the geometric separation processing [179].

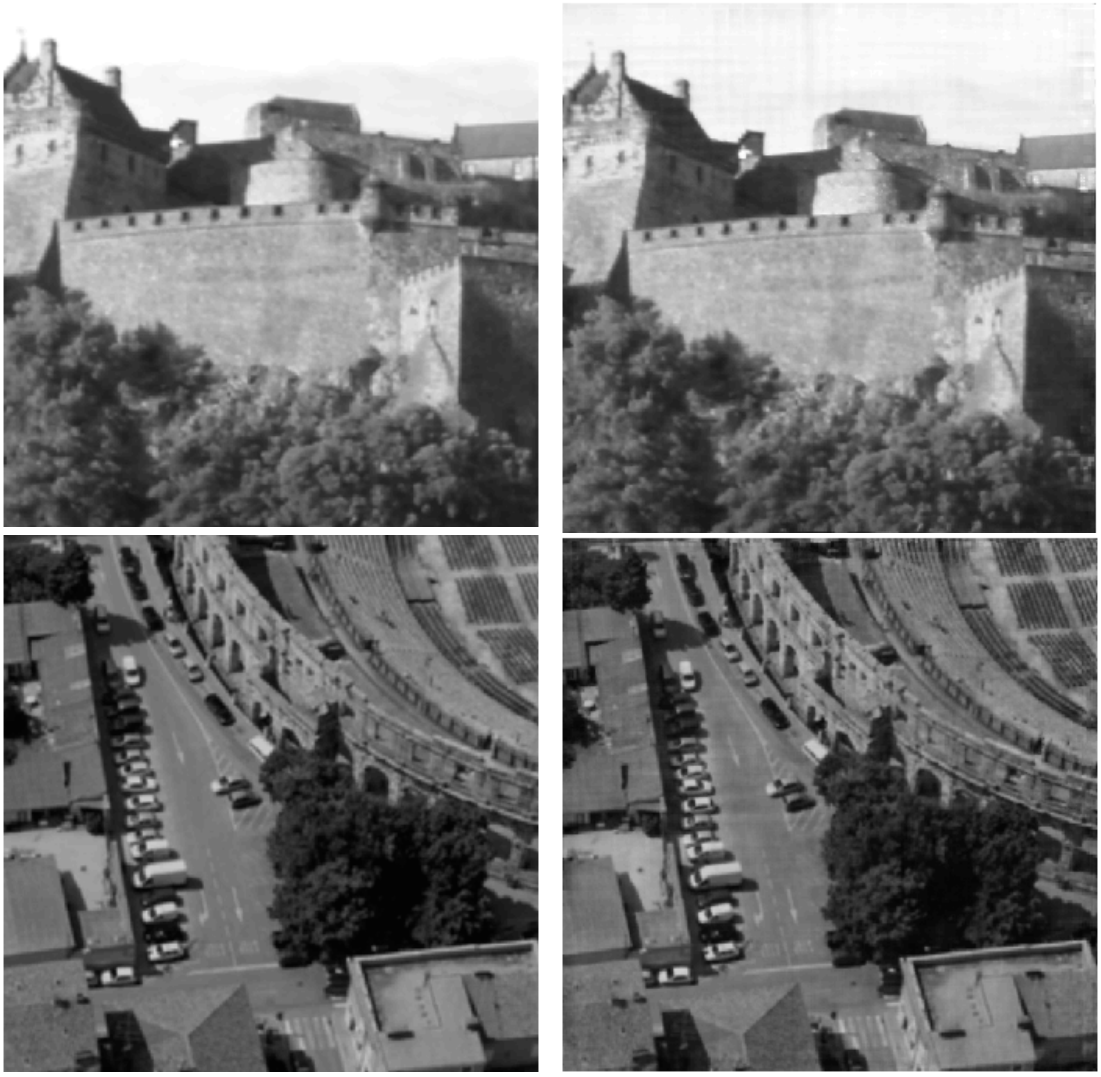


Figure 6.9: $nFoV$ sections of super-resolution image from 50 images (left) and filtered separated $nFoV$ images (right).

Figure 6.6 and Figure 6.7 show the findings for sample image [B]. Figure 6.6 shows the averaged pixel intensity error percentage and Figure 6.7 shows the relationship between the averaged SSIM index and the number of frames. It can be seen that once again the super-resolution method {1}, initially provides the best image quality, although marginally as illustrated by Figure 6.7, but within $5-10$ frames the improvements from both methods plateau, providing very little new information.

Figures 6.8-6.10 display the resultant input and output images. Figure 6.8 shows the initial results from a single frame for the $nFoV$ images. The blocky nature of the noise introduced into the images is characteristic of the geometric separation technique [179]. Figure 6.9 shows the results after the $50th$ frame for both test images, for the filtered geometric separation technique and the close-up of the $nFoV$ region in the super-resolved image. Figure 6.10 shows the filtered $wFoV$ {4} and the full $wFoV$ super-

resolved image for comparison.

Figure 6.10: $nFoV$ sections of super-resolution image from 50 images (left) and filtered separated $nFoV$ images (right).

The purpose of the $nFoV$ for image sensor applications is to provide the high spatial details required for the interrogation of a region-of-interest or target. Figure 6.9 provides a side-by-side comparison of the $nFoV$ images, I_2, obtained by the time-averaged geometric separation and the super-resolved image, I_1. Qualitatively both images look very similar to each other. Quantitatively the super-resolution method is the same, or marginally better, as shown in Figures 6.4-6.5 and Figures 6.6-6.7 with a structural similarity of around 90%. The filtered $nFoV$ images remove nearly all of the block noise/artefacts that are introduced by the geometric separation method [179] applied to a single frame (see Figure 6.8) and it provides a performance level similar to that of the super-resolution technique, at least as measured by the structural similarity measure. Looking at the vehicles present in the aerial image in Figure 6.9,

both the geometric separated image and the super-resolution image provide sufficient visual structure/details to distinguish different vehicle types in the processed image.

6.4.2 Sensitivity to Image Registration Errors

In order to determine the robustness of each technique it is important to understand how inaccuracies in image registration affect the performance. The SSIM index and the averaged pixel intensity error values were used once again, to evaluate the findings. The average image registration error was varied by a tenth of a pixel in the horizontal and vertical directions for each interval. The averaged pixel intensity error and SSIM values shown in Figures 6.11-6.14 have been averaged over 50 individual realisations, each realisation consisting of 50 frames.

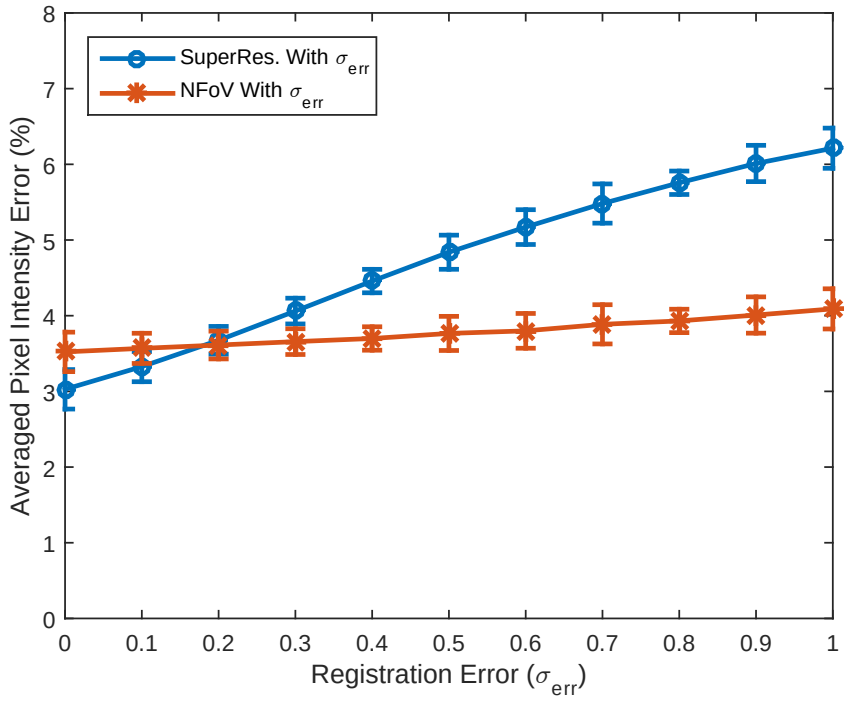


Figure 6.11: Averaged pixel intensity error calculated for various image registration errors for the time-averaged $nFoV$ and SR methods using sample image [A]. Each registration error value consists of 50 runs, each run containing 50 frames, with error bars (1 standard deviation).

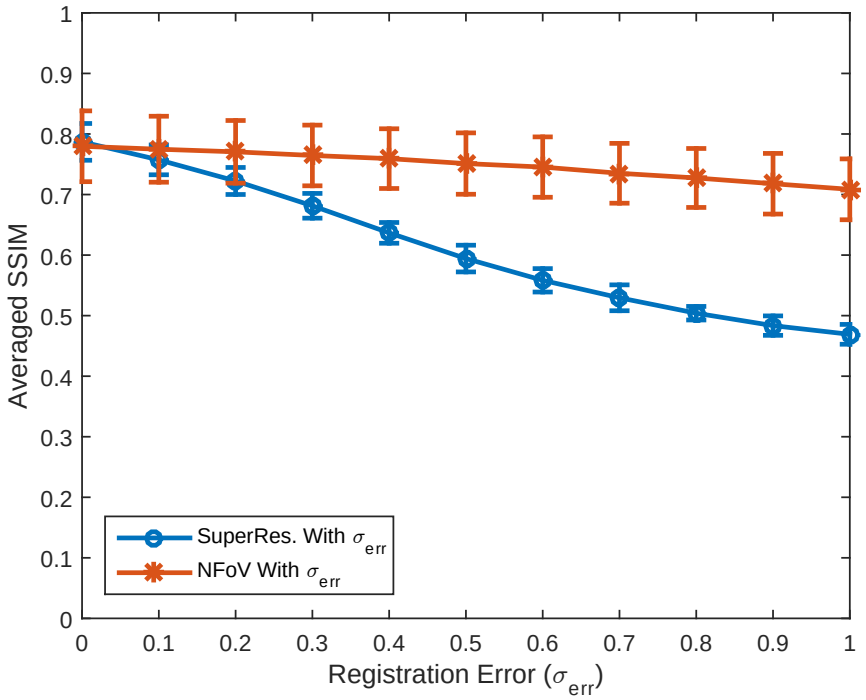


Figure 6.12: SSIM calculated for various image registration errors for the time-averaged $nFoV$ and SR methods using sample image [A]. Each registration error value consists of 50 runs, each run containing 50 frames, with error bars (1 standard deviation).

Figure 6.11 and Figure 6.13 show that when there is perfect registration between frames, the SR method outperforms the time-averaged method, but by only a relatively small margin. As the registration error increases, it can be seen that both methods begin to deteriorate in terms of image quality. However, the image quality of the super-resolution technique deteriorates rapidly compared to the time-averaged technique.

Figure 6.12 and Figure 6.14 show that with perfect registration, both the super-resolution method and the time-averaging method produce images of similar quality as measured by SSIM, but as the image registration error increases the super-resolution method begins to deteriorate rapidly, whereas the time-averaged method only suffers a loss of approximately 10% overall.

The figures show that whilst both methods suffer image quality loss as the registration errors increase, it can be seen that the time-averaged geometric separation technique is much more robust to the variations in registration error. Even with perfect image registration, super-resolution only marginally outperforms or produces similar quality images compared to the time-averaged geometric separation technique.

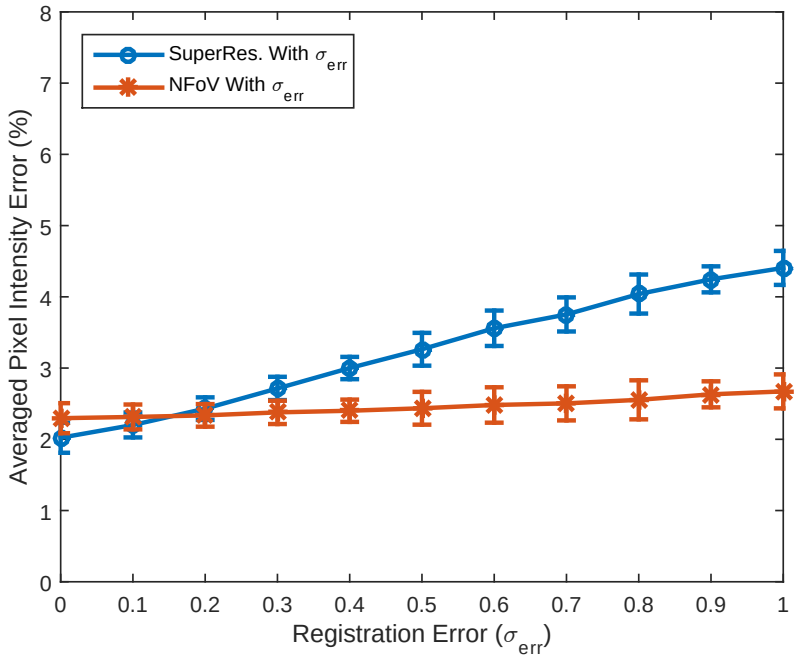


Figure 6.13: Averaged pixel intensity error calculated for various image registration errors for the time-averaged $nFoV$ and SR methods using sample image [B]. Each registration error value consists of 50 runs, each run containing 50 frames, with error bars (1 standard deviation).

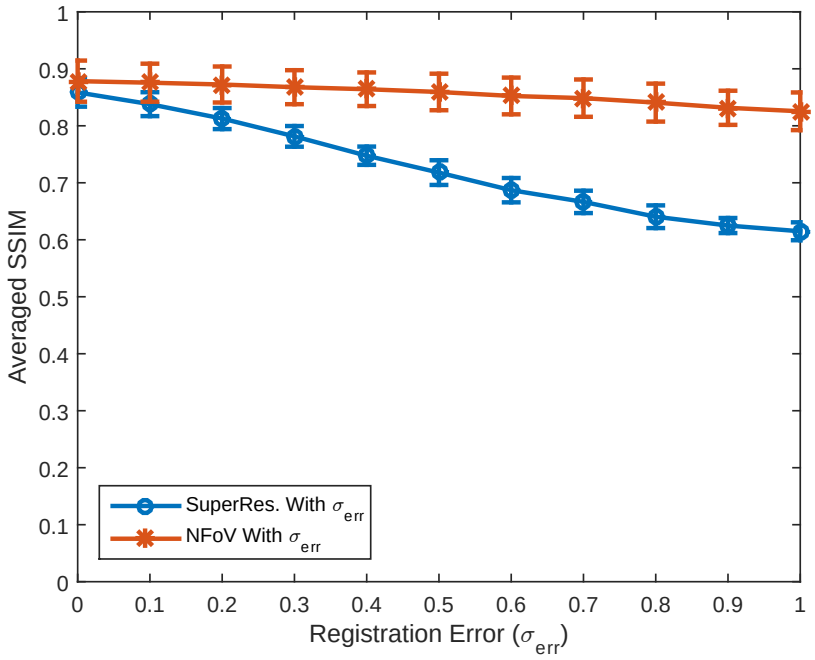


Figure 6.14: SSIM calculated for various image registration errors for the time-averaged $nFoV$ and SR methods using sample image [B]. Each registration error value consists of 50 runs, each run containing 50 frames, with error bars (1 standard deviation).

The initial differences between the averaged pixel intensity error and the SSIM for small registration errors could be due to the different image characteristics each image metric includes in its assessment of image quality. The averaged pixel intensity error only considers the individual pixel contrast values and calculates the mean percentage error. On the other hand the SSIM index not only considers contrast, but also luminance and structural information, providing a more accurate comparison of the images, using a broader spectrum of image properties [276].

6.5 Chapter Summary

This chapter has proposed an image separation technique for use with the novel imaging system for use by aircraft such as, missile systems and uninhabited aerial vehicles (UAVs). The imaging system superimposes two images; one wide field-of-view and one (co-boresighted) narrow field-of-view of the same scene. Previous work, in Chapter 4, introduced a geometric separation technique that provided a post-processing recursive algorithm to separate the superimposed image into its constituent $wFoV$ and $nFoV$ images [179]. The initial algorithm demonstrated that single frame, superimposed images could be separated but the resultant separated component images suffered from macro-blocking artefacts that limited the image quality. The method proposed in this chapter has extended the single-frame geometric separation technique by implementing a time-average filter, which enables the technique to work over multiple frames (or video streams). The quality of the resultant images has been assessed quantitatively using the averaged pixel intensity error and the SSIM index. It has been demonstrated that the geometric separation method provides an image quality that approaches that found using an optimal super-resolution technique. The super-resolution provides better quality images for perfectly aligned frames, but with the introduction of image registration errors the quality of the super-resolution images degrades rapidly. The time-averaging geometric separation technique provides a more robust solution which can be critical in certain applications such as tracking and surveillance.

The principal advantage of using superimposed images and a time-averaged geometric separation technique rather than a conventional imaging system for aircraft applications is that it can provide both wide field-of-view situational awareness and surveillance alongside high resolution imagery for object identification and tracking using a single imaging system.

Chapter 7

Image Quality Assessment

7.1 Introduction

In this chapter, the work presented in Chapter 6 will be extended over multiple images from a variety of acquisition methods, image formats, compression techniques, exposure, brightness and image content (i.e. structural information). A database of images has been collated and presented in Appendix B for this study. The results in Chapter 6, were limited to two images and evaluated using only the averaged pixel intensity error and structural similarity index (SSIM), as image quality metrics (IQMs).

In this chapter, a review of the different image quality assessment (IQA) metrics available will be conducted, a selection of the IQMs which provide the best quality assessment will then be chosen from a range of different types of IQM methods. A new IQM will also be proposed, called the image quality information metric (IQIM), and used in conjunction with the selected IQM techniques. The temporal geometric separation technique and super-resolution techniques will then be performed on the entire image database, following the work described in Chapter 6 over varying registration errors. The work presented in this chapter aims to provide a broad scope analysis of the research carried out in the previous chapter, for a large variety of images and registration errors.

Image quality assessment (IQA) is the estimation of an image's perceived quality, and its deviation from the original (ideal) image. IQA algorithms aim to consistently imitate human quality perception, to provide a reliable estimate of image quality [218]. The quality of an image represents the amount of visual degradation present and can be used for many image processing applications (i.e. medical imaging). Image degradation is defined as the reduction of the inherent optimum potential of an image and its sensor system, due to errors in sensor operations, processing procedures or incorrect image handling [199]. Degradation/distortion can be introduced into an image due to loss/corruption of image information, during image acquisition, compression, storage, transmission, decompression, display, or even printing phases [141]. Degradation can be present in many forms such as noise, blocking artefacts, blurring and fading [99,141].

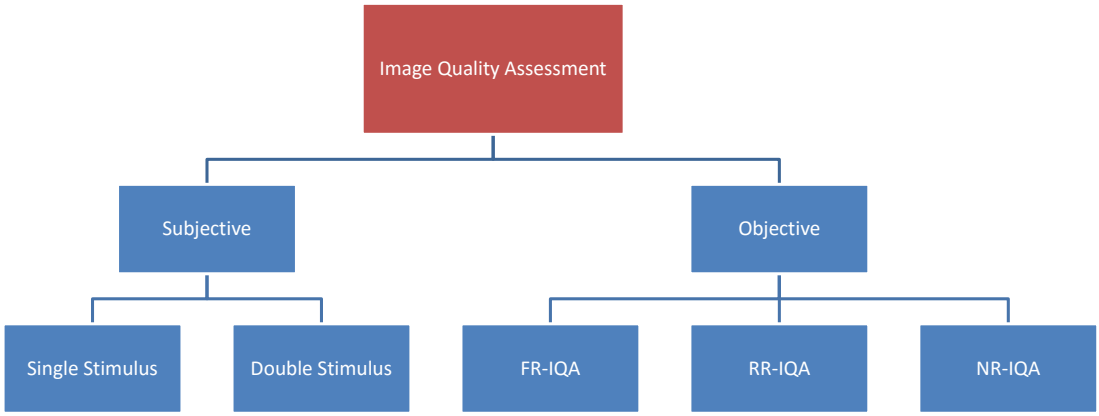


Figure 7.1: Overview of the different types of image quality assessment methods

7.2 Subjective IQA

Image quality can be measured in two main ways, subjective and objective methods. In subjective methods, human subjects are utilised to assess the visual quality. The human vision system (HVS) can easily identify forms of distortions [141] that lead to degradation and generally functions by extracting structural information from the images [276]. The inherent problem with using the HVS is the subjectivity itself. Each individual's opinion is based on their own experiences and training, hence there can be large variation between individuals. Subjective methods utilise the HVS for assessing image quality but generally the methods can be expensive and time consuming.

The different Subjective IQA Methods used are:

1. Single Stimulus Method
   In the single stimulus method a set of stimuli (images) are presented, one at a time for a fixed duration [170]. The observers are then required to rate the images into one of five categories: excellent, good, fair, poor or bad or alternatively as a numerical category rating. Only a single judgement is required per assessment [208].

2. Quality Ruler Method
   The quality ruler method utilises a set of reference images of known quality that are evenly distributed along a pre-calibrated quality scale (Standard Quality Scale). The assessed quality scores for new images then correspond to their position along the scale [99].

3. Mean Opinion Score
   The mean opinion score method is a subjective measurement, which calculate the arithmetic mean, of all individual scores, provided by the assessors/participants. The assessor is required to give a quality rating from 1(Very poor) to 5(Excellent).

An example of where this methodology was used in the telephony network, which used mean opinion score to assess their VoIP (Voice over Internet Protocol) calls quality [53].

Table 7.1: Mean Opinion Score Table

1	2	3	4	5
Very Poor Quality	Poor Quality	Good Quality	Very Good Quality	Excellent Quality

The single stimulus method is often chosen for subjective testing, as it is straightforward to implement and well standardised [218]. There are some drawbacks of this method, the major disadvantage being that the spread between different assessors can be relatively large. From the selection of observers, their knowledge, expertise, availability, seriousness, bias interpretations and other various factors can influence their decision [141].

Subjective testing has the potential to be the most reliable methodology, as the HVS is used directly for quality perception, but can be there are large variations in opinions due to experience and training of an individual, it can also be expensive and time-consuming. Therefore, there is high demand for objective IQA methods, which can provide automatic quality prediction by imitating the human visual system (HVS).

7.3 Objective IQA

The aim of research into Objective IQA metrics is to be able to predict the perceived quality of an image using mathematical expressions to remove any bias, as experienced with subjective methods. There are many applications that can benefit from the use of objective IQA methods, for example [258]:

1. An image quality metric (IQM) can be used as a benchmarking tool, to evaluate the performance of image and video processing systems and algorithms, such as compression and image restoration techniques. A reliable IQM can help determine the best compression method available, which will provide the least amount of image/video degradation.

2. An IQM can help monitor image quality in quality control systems. In video streaming applications, IQMs can help monitor the quality of a transmitted video signal, in order to transmit the optimum quality signal to its receiver(s).

3. IQMs can be used to optimise video and image processing systems. They can be used to calibrate parameter settings for encoders/decoders in order to optimise the reconstruction of a video stream.

Objective image quality assessment can be classified into three categories, dependant on the availability of the reference image. The reference image is the ideal/original image. The three categories are [141]:

1. Full reference image quality assessment (FR-IQA)
2. Reduced reference image quality assessment (RR-IQA)
3. No reference image quality assessment (NR-IQA)

7.3.1 Full Reference IQA

Full reference image quality assessment (FR-IQA) refers to quality perception methods which utilise the reference image to assess the quality of the distorted image [235]. The reference image would ideally be the undistorted, original source image of the distorted version. A standard measurement used to assess image quality is the mean-square error (MSE), defined by [275, 278]:

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - y_{ij})^2 \tag{7.1}$$

For a given monochrome $M \times N$ image x and its noisy approximation y. One problem with MSE is that it is strongly dependent on the image intensity scaling [104]. Another alternate method for evaluating image quality is the peak signal-to-noise ratio (PSNR), defined as [128, 134]:

$$PSNR = -10 \log_{10} \frac{MSE}{S^2} \tag{7.2}$$

Where S is the maximum possible pixel value of the image and MSE is the mean-spuared error. However it has been shown that PSNR does not correlate consistently with human visual perception and image quality [128]. Other mathematical distortion measures such as the average difference, absolute error, Mean MSE, Laplacian MSE and many others are described in [80].

In order to improve on traditional mathematical metrics such as PSNR and MSE, other parameters have been included in the development of image quality metrics which were perceived to correlate with the human visual system (HVS). In HVS based image quality metric models the difference between the reference image and the distorted image is normalised according to its perceived visibility of the HVS. The main features that are commonly used in HVS IQM are:

Contrast Sensitivity Function (CSF). Contrast is the difference in luminance of an object and it background. Contrast sensitivity is the ability to distinguish an object from its background. CSF models the variation in the sensitivity of the

HVS to different contrast thresholds [155]. Some IQM models implement this using a Low-pass (or band pass) filter, others implement a weighting factors for subbands after frequency decomposition [277]. The contrast sensitivity is also a function of temporal frequency, but it is irrelevant for image quality assessment. It has been modelled in video quality assessment as a temporal filter [265,266,281].

Luminance Contrast Sensitivity, also called "Luminance Masking". Human eyes are sensitive to luminance contrast (i.e. the change/difference) in luminance, rather than absolute luminance value [258]. The luminance contrast can be defined as a ratio of the luminance difference (ΔL) and the average luminance (L_{AVG}), $\Delta L / L_{AVG}$, derived from the Weber-Fechner law. If the average luminance contrast is high, a small luminance difference is negligible, but if the average luminance was low, then it would become relevant. Some IQM models implement this as a detection threshold [98].

SSIM

As used previously in chapters 4 and 6, the structural similarity (SSIM) index [275,276], is regarded as one of the most highly correlated metrics to the HVS, and robust FR-IQA metric available. It is defined as a function of luminance, contrast and structural information.

In this approach, it is assumed that the HVS is highly adapted to extracting information from natural scenes, which are highly structured. Therefore, measuring the structural information change between the reference and distorted image should provide a good estimation of the perceived image quality. The SSIM index is defined as [276]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{({\mu_x}^2 + {\mu_y}^2 + C_1)({\sigma_x}^2 + {\sigma_y}^2 + C_2)} \tag{7.3}$$

where μ_x and μ_y are the mean intensities in the source image signal x and the distorted image signal y which help describe the luminance term. σ_x and σ_y are the source and distorted image signals standard deviation, used to describe the contrast comparison. C_1 and C_2 are constants to avoid instability when $\mu_x^2 + \mu_y^2$ tends towards zero and $\sigma_x y$ is the correlation co-efficient corresponding to the cosine of the angle between vectors $(\mathbf{x} - \mu_x)$ and $(\mathbf{y} - \mu_y)$, providing a measure for the structural term [276].

MS-SSIM

Multi-scale SSIM (MS-SSIM) IQM is an extension from the previous SSIM IQM [280]. The MS-SSIM is designed to be more flexible than the single-scale (SSIM) method by incorporating the variations of viewing conditions such as display resolution, viewing orientation and viewing distance. Using the reference and distorted images as signal inputs, a low pass filter is applied and the filtered images are then downsampled by a

factor of 2. The original image is indexed as scale 1 and the highest scale is referenced as scale M. The contrast comparison and structure comparison are calculated at each scale. The luminance is computed at scale M only. The overall metric is obtained by combining the measurements from each scale, which are weighted dependent on the contrast sensitivity function (CSF) using a image synthesis approach for cross-scale calibration [277].

Fast SSIM

Fast SSIM (F-SSIM) [47] was developed using the previous knowledge of SSIM and MS-SSIM but with particular focus on real-time applications. F-SSIM modifies the luminance and contrast terms previously developed for the aforementioned image metrics, in order to optimise the computational speed and reduce computational complexity. Parallel computing and sub-sampling is also employed in order to extract the best possible performance for real-time applications [47].

Visual Information Fidelity

The visual information fidelity (VIF) [236,237] criterion is established on the initial paper of the information fidelity criterion using natural scene statistics [238], and a simple human visual system error model. VIF is defined as the ratio of mutual information between the original and distorted images to the information content of the original image.

7.3.2 No Reference IQA

No reference image quality assessment (NR-IQA), also referred to as blind IQA, is the case when the original undistorted source image is not available and no information of its properties are available/obtainable. At present, NR-IQA methods can be categorised into two sub-categories: Distortion-specific and distortion-generic methods. Distortion-specific methods tend to focus on assessing the image quality based on certain distortion effects regardless of other factors, whereas distortion-generic methods aim to predict the image quality by attempting to detect abnormal/un-natural features, thought to be due to distortion using a training technique such as support vector machines (SVM). Some examples of NR-IQA are outlined below.

DIIVINE

Distortion identification-based image verity and integrity evaluation (DIIVINE) [192] is based on the two-stage framework; distortion identification followed by quality prediction. In the first stage, natural scene statistics are used to to extract and identify distorted features. In the second stage, the image quality is predicted by taking the

selected features in the wavelet domain to calculate the wavelet co-efficients and using the Gaussian scale mixture (GSM) model presented in [270] to predict image quality.

BRISQUE

The blind image spatial quality evaluator (BRISQUE) [189] is a natural scene statistic-based (NSS) distortion-generic, spatial domain NR-IQA technique. Many NR-IQA techniques assess image quality by calculating distortion specific features, for example, features with blurring, blocking or ring effects. BRISQUE utilises the scene statistics based on the distribution of locally normalised luminance coefficients and uses a statistical approach to quantify the possible losses of naturalness due to the distortion effects.

BLIINDS-II

Blind image integrity notator using DCT statistics (formally known as BLIINDS-II) is the advancement of the former (BLIINDS-I) [227]. The former method was a single-stage framework which used no statistical modelling. BLIINDS-I predicts the quality of an image by analysing the statistics of the local discrete cosine transform (DCT) coefficients. BLIINDS-II is a two-stage framework which uses a NSS model in the DCT domain, the local DCT coefficients are used for feature extraction. The NSS features are then used by the Bayesian probabilistic model to predict the image quality [228].

SSEQ

Spatial-spectral entropy-based quality (SSEQ) [162] index is a distortion-generic NR-IQA technique which uses a two-stage framework of distortion classification followed by a quality assessment stage. SSEQ uses a support vector machine (SVM) to train an image distortion and quality prediction engine. It is important to note that the quality prediction stage is dependent on the training database and the variety of distortions present in the training set.

7.3.3 Reduced Reference IQA

Reduced Reference IQA (RR-IQA) methods are used when some information of the original source image is known, and can be used to assess the quality of the distorted image. In general, RR-IQA metrics tend to be modified versions of FR-IQA techniques, which have limited access to the source image information. For example, a reduced reference variant of the SSIM index has been proposed by [219]. Other RR-IQA schemes such as wavelet-domain methods using natural scene statistics have also been developed [279]. RR-IQA methods are out of the scope of this thesis and therefore will not be discussed further, because the aim of the study is to assess the quality of the separated images, to the best of our ability and therefore it is better to use the FR-IQA methods

described previously, as RR-IQA methods tend to be modified variants of FR-IQA metrics, and because the source images are available for our comparison purposes.

7.4 Image Quality Information Metric

In this section, a new image quality metric (IQM) is introduced. The image quality information (IQIM) metric takes a number of $n \times n$ patches (m patches, each of $n \times n$ pixels) from the source image. Each patch is transformed to the discrete cosine transform (DCT) domain. The discrete cosine transform represents a finite sequence of data points as the sum of cosine functions oscillating at different frequencies [3, 215].

The DCT transform generates a real $n \times n$ matrix of spatial frequencies corresponding to the selected $n \times n$ image patch [167] and a corresponding $n \times n$ matrix of transform coefficients. The linear combination of the spatial frequencies (DCT basis functions) and the transform coefficients would reproduce the image.

Figure 7.2 shows the DCT basis functions; a combination of horizontal and vertical frequencies for an 8×8 two-dimensional DCT patch. Following the description from [215], each step from left to right and top to bottom represents a step increment, in frequency by a half-cycle. For instance, moving right by one from the top-left DCT basis function (top-left square) yields a half-cycle increase in the horizontal frequency. Another move to the right amounts to, two half-cycles (i.e. one cycle in total). A move down, from the previous position, results in two half-cycles horizontally and a half-cycle vertically. The top left frequency component on the basis function is the DC (Direct Current) or the zero spatial frequency component. The source data is transformed to a linear combination of these 64 frequency squares. The weightings that transform the basis function into the source image are known as the transform coefficients. The transform coefficients are multiplied by it respective DCT Basis component and then summed together to recreate the source data/image.

Operating in the DCT domain offers certain advantages. One such advantage in the DCT domain is there evidence to suggest that the DCT information is closely correlated to the human visual system (HVS) [78, 142, 163, 168, 300]. In the field of psycho-visual encoding, 8×8 DCT domain patches are utilised in JPEG psycho-visual encoding methods [31]. Psycho-visual encoding is the process by which the human visual system constructs and stores images in the human brain. In digital data compression techniques, psycho-visual redundancy is redundant data that can be removed without substantially affecting the data/image signals as interpreted by the HVS [240].

Assuming that the DCT components are independent of each other, the IQIM takes the DCT transform of the m, $n \times n$ patches and creates a probability distribution function (pdf) using a histogram 'binning' process, for each of the DCT components (i.e. $n \times n$) (see Figure 7.3). This provides an estimate of the pdf for each of the individual DCT components (i.e. 64 ($n \times n$)). See Equation 7.5.

In DCT coding, each component of the image is sub-divided into block of 8 pixels (for JPEG compression and most other uses). A 2D DCT is applied to each block of data to obtain and 8×8 array of co-efficients. Taking a block of $x(n, p)$ to represent the image pixel values in a block (where $n = p = 8$), the DCT is computed using:

$$X(u, v) = \frac{c(u)c(v)}{4} \sum_{n=0}^{7} \sum_{p=0}^{7} x(n, p) \cos \frac{(2n + 1)u\pi}{16} \cos \frac{(2p + 1)v\pi}{16} \tag{7.4}$$

For: $0 \leq u \leq 7$ where: $c(u) = \frac{1}{\sqrt{2}}$ when $u = 0$

$0 \leq v \leq 7$ $c(u) = 1$ when $1 \leq u \leq 7$

where, $X(u, v)$ is the DCT coefficient at coordinates (u, v) and $c(u)$ is a normalising scale factor to make the transformation orthonormal. u and v are the horizontal and vertical spatial frequencies.

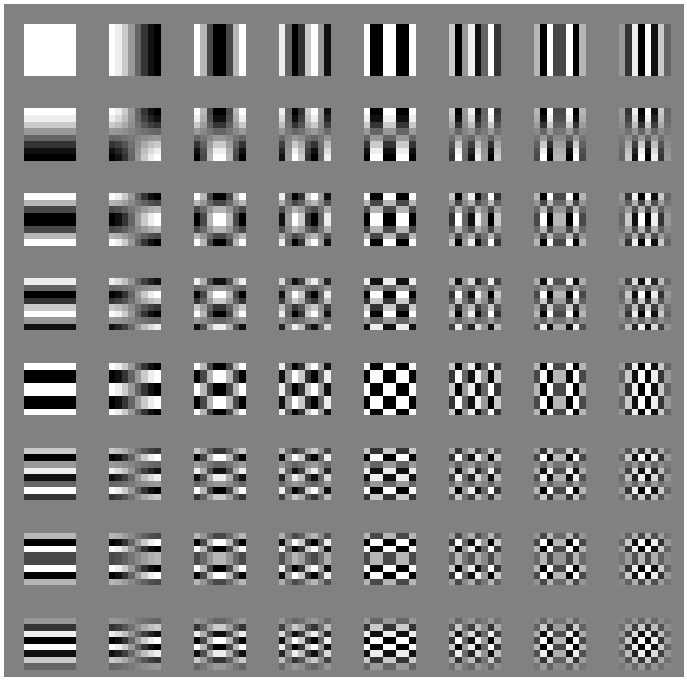


Figure 7.2: Two-dimensional DCT basis function generated from [173]. The image shows the combination of different horizontal and vertical frequency combinations for an 8×8 DCT. Each step, starting from the upper left corner in the horizontal or vertical direction is a half-cycle increment, in the respective horizontal or vertical frequency. The source data is transformed from the spatial domain to the DCT domain and is represented as a linear combination of the (8) 64 frequency components (squares).

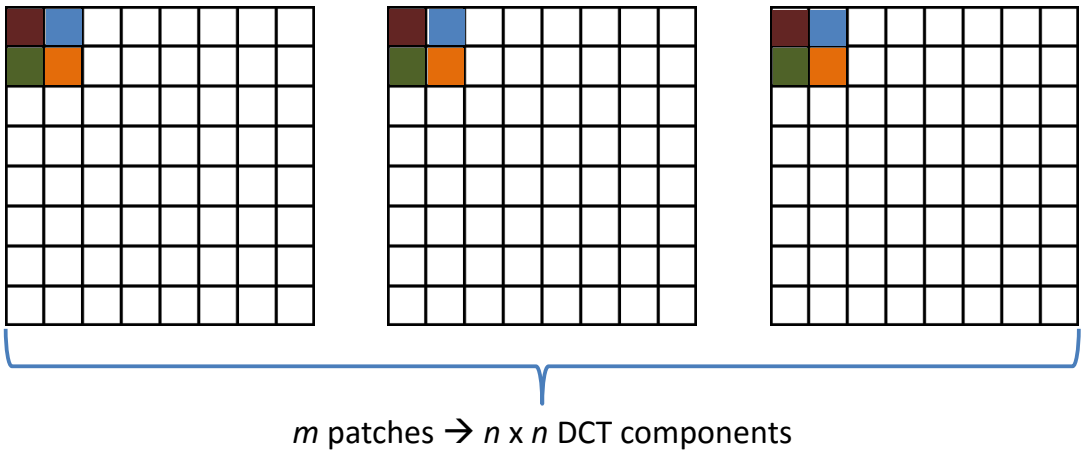


Figure 7.3: Illustration of m DCT patches, with each DCT component (i.e. red, blue, green or orange DCT components etc.) cumulatively summed to create 64 histograms (i.e. 8×8 individual DCT components) over all m patches.

$$I_{(i)} = -\sum_{i=1}^{n \times n = N_{tot}} p_i \log_2 p_i \quad (bits) \tag{7.5}$$

where $I_{(i)}$ is the information for the i^{th} DCT component over all m patches. This will result in a total of 64 $(n \times n)$ information values, one for each DCT component over the stack of m patches, as the histograms are accumulated over all m patches. The information is calculated based on the Shannon information formula [234].

The 64 information values are summed together to calculate the maximum possible information across an 8×8 DCT patch. This is the maximum information possible because each DCT component of the 8×8 DCT patch (or 64 components) is assumed to be independent from each other. This is because the correlation between neighbouring DCT components will reduce the amount of information available.

From this, the sum of all the DCT components in the m patches is used to calculate the estimated maximum information present using the Shannon information (Equation 7.6). It is important to note that the number of patches selected can affect the results. If there are not enough patches, the estimate of the pdf may not fully represent the image, but conversely too many patches and this will affect the computational time. A study to show the optimal number of patches will be conducted in Section 7.4.2.

$$I_{max} = \sum_{i=1}^{n \times n = N_{tot}} I_{(i)} \quad (bits) \tag{7.6}$$

where I_{max} is an estimate of the maximum information available in m patches and p_i is the estimated probability for the i^{th} DCT component, obtained from the pdf histograms.

Extending this, the estimated information can be improved by estimating the mu-

tual information between neighbouring DCT components. This is because the assumption of independent DCT components is highly unlikely, the majority of correlation between DCT components would be found in the neighbouring horizontal, vertical and diagonal DCT components (see Figure 7.4). To obtain this, the correlated information between neighbouring DCT components within a 8×8 DCT patch are calculated using the joint probability distribution function (joint pdf). This gives a measure of correlation between the components (i.e. mutual information). This can be calculated for the adjacent vertical, horizontal and diagonal elements (see Figure 7.4).

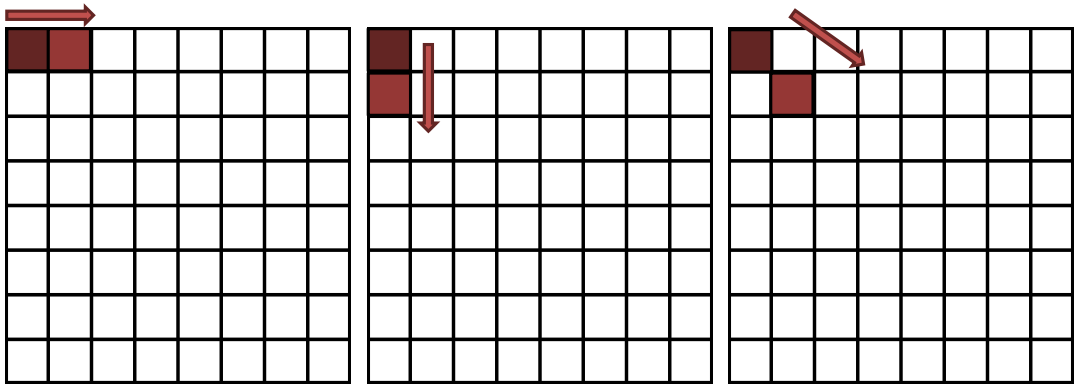


Figure 7.4: Illustration of horizontal, vertical and diagonal neighbouring components in a DCT patch

The joint pdf is constructed in a similar manner to the pdf histograms described above, but instead of a single variant histogram, an bivariate (2D) histogram is produced, because the probabilities are based on joint occurrences (i,j) instead of just (i) as in the previous method (see Figure 7.5).

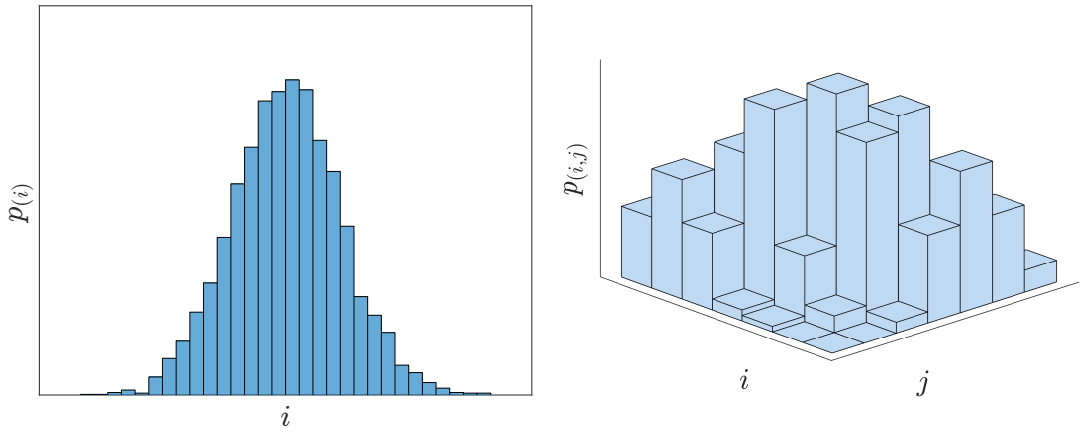


Figure 7.5: Illustration of single-variate (1D) (Left) and bivariate (2D) (Right) histograms

$$I_{mut}^{(h/v/d)} = -\sum_{i=1}^{n\times n}\sum_{j=1}^{n\times n} p_{(i,j)} \log_2 \frac{p_{(i,j)}}{p_{(i)}p_{(j)}} \quad (bits) \tag{7.7}$$

where $I_{mut}^{(h/v/d)}$ would be the mutual information in either the horizontal, vertical or diagonal DCT components in a 8×8 DCT patch.

$$I_{mut} = I_{mut}^{(h)} + I_{mut}^{(v)} + I_{mut}^{(d)} \tag{7.8}$$

The summation of the horizontal, vertical and diagonal mutual information of all DCT components for all m, $n \times n$ patches gives us the mutual information between all elements for all m DCT patches. Thus by subtracting the mutual information between DCT components (Equation 7.8) from the maximum information available (equation 7.6) yields a better estimation of the information between the m patches (Equation 7.9).

$$I = I_{max} - I_{mut} \tag{7.9}$$

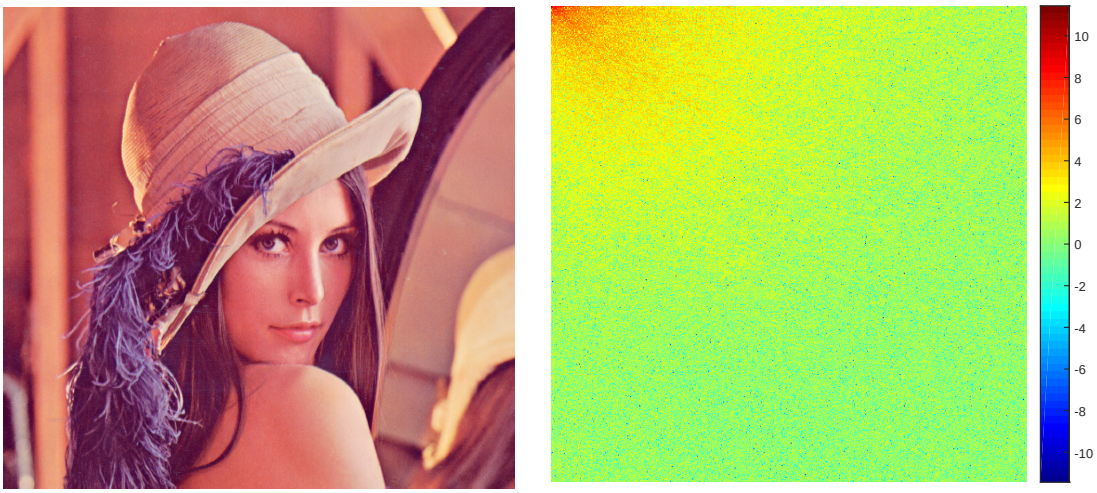


Figure 7.6: Lena image (Left) with the corresponding DCT domain (Right) representation the top-left corner shows the zero spatial frequency (DC) component and the bottom left show the high spatial frequency component

In the field of psycho-visual encoding, the utilisation of 8×8 sized patches is the standard and has been used widely in encoding techniques such as JPEG [54]. Therefore, the weightings are readily available, and would be a suitable standard to use for the 8×8 sized patches. The JPEG standard first began development in 1986 and the first JPEG standard was created in 1992 and is under ISO/IEC 10918 standards named *Information technology Digital compression and coding of continuous-tone still images* [2]. The JPEG standard is currently composed of 6 parts, as of 2012 [54, 135].

Part 1: Specifies the core coding technology and options for encoding photographic images

Part 2: Defines the testing for compliance

Part 3: Defines a set of extensions to the coding technologies within Part 1.

Part 4: Focuses on the registration of JPEG profiles, still picture interchange file format (SPIFF) profiles, SPIFF tags, SPIFF colour spaces, SPIFF compression types, and defines the registration authorities

Part 5: Specifies the JPEG file interchange format (JFIF)

Part 6: Focuses on tools for application to printing systems

The JPEG encoding technique uses the DCT frequency-domain. More specifically two-dimensional type-II DCT and its corresponding inverse (IDCT) type-III DCT transform as described by [215]. One of the aims of the JPEG codec is to remove psycho-visual redundant data, as it is unnecessary to keep this, because the HVS cannot process this information and therefore it takes up non-essential storage capacity. As part of the IQIM we will calculate the information and mutual information for the JPEG equivalent images as this will allow to compare the metric to the JPEG standard which is designed to be tailored for the HVS and will serve as a good comparative measure, in order to assess the suitability and accuracy of the IQIM metric and its correlation to the HVS.

$$JPEG_{mat} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \tag{7.10}$$

7.4.1 Methodology & Implementation

The method above relies on selecting a good sample set of image patches to best represent the information in the entire image. The two most widely used sampling methods available are random sampling and uniform sampling. It is also important to select a large enough sample set to produce stable and reliable results over several different image types whilst also minimising excessive computational run-time costs.

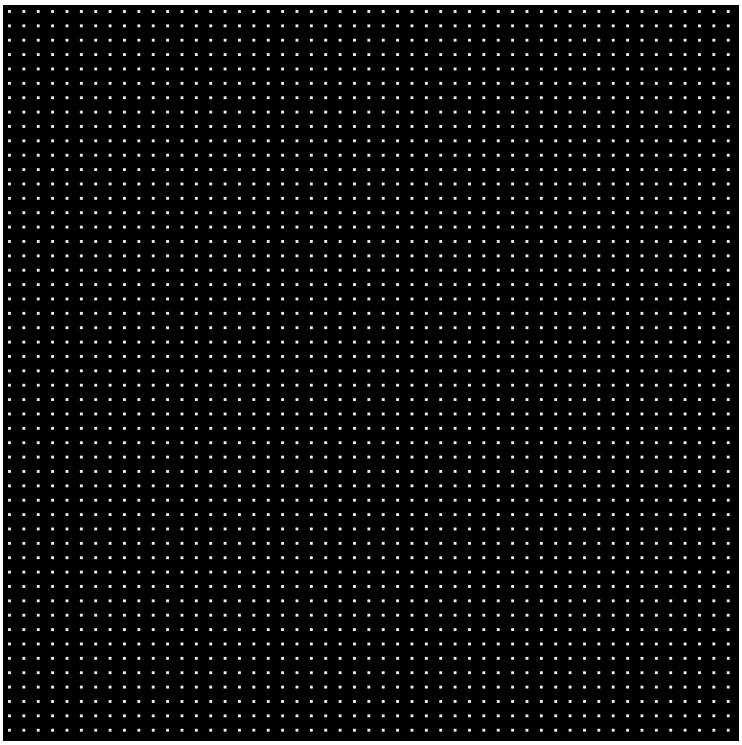


Figure 7.7: Uniform sampling illustration with 2600 sample points

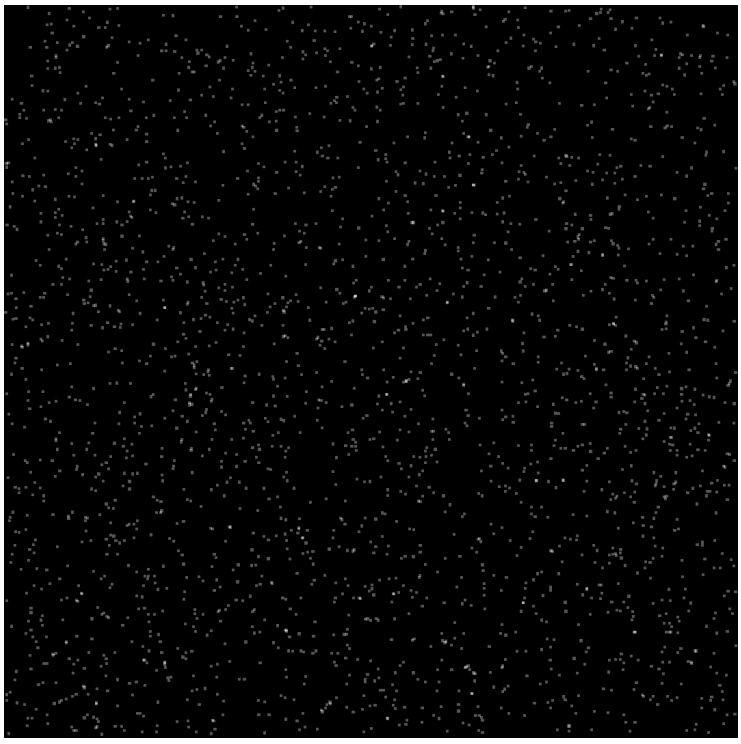


Figure 7.8: Random sampling illustration with 2600 sample points

A case study to evaluate the best possible sampling method and the optimal number of patches required was performed, allowing to generate reliable results without unnecessarily large computational overheads. To perform this test, random sampling and uniform sampling patch selection codes were developed using Matlab® [177]. The Matlab in-built random number generator was used to select the ordinates of the patches for the random sampling method. The Matlab random number generator was reset every time the code was initialised using the current date and time stamp acquired from the machine's operating system. Once the m, $n \times n$ image patches have been extracted, the patches are transformed to the DCT domain and JPEG weighted versions are also made. From this, the maximum information (Shannon entropy) (7.5) and mutual information (7.8) are calculated, and finally the actual information (7.9) is estimated.

Matlab Performance Considerations

The storage and processing of large images can be computationally expensive, therefore it is vital to optimise the code to reduce memory consumption and computational run-time costs. The DCT patches are stored as stacked array of dimension $(n \times n \times m)$. Structuring the DCT patches in this form enables the use of simple arithmetic operations to be calculated over the entire stacked array rather than individual patches. This helps improve the computational efficiency of the program. Furthermore, using stacked arrays allows the user to implement Matlab's Parallel Computing Toolbox™, enabling the use of multiple CPU threads at the same time. Another possible optimisation technique is the use of sparse matrices, but due to the large overheads of restructuring and re-indexing of the stacked array, it was found to be inefficient in our application and therefore removed after testing. Memory management is very important when dealing with large arrays. The pre-allocation and clearing of non-essential memory is key for optimal performance without sacrificing stability and the reliability of the results.

7.4.2 Uniform vs Random Patch Sampling

Determination of the optimal sampling method is crucial in the patch selection process in order to prevent inaccurate and unreliable measurements. An important parameter to be considered in the evaluation is the number of patches required for both sampling methods, which allows for reliable results with different image types.

Images can be acquired from many different sources and go through various stages from acquisition, processing to storage and display/printing. Due to this, the image quality can be degraded by the many factors and processes that the images are subjected to, for example, compression for storage purposes. The types of images/scenes being captured, also play a key role in the creation of the image and its "image life-cycle". Therefore, to provide accurate results for all images types, the patch sampling test

has been performed over a small subset of images from the image database (Appendix B) varying from different contrast ranges, exposure, landscapes, vector images, raster images, highly structured images, blur, image artefacts, image distortion.

The results of the proposed IQIM technique will be compared with the SSIM index [276], which is arguably considered one of the most widely used full-reference (FR-IQA) image quality tools available [68]. An image database has been created to collate high resolution imagery from various sources, containing different image properties, to generated the best possible results. Current publicly available image databases such as LIVE [239], Microsoft© [254] and others listed in [123,132,255] contain images that are not at a resolution high enough for the purposes of our simulations. Therefore, a small database of high resolution imagery with a minimum resolution of 2048 × 2048 pixels was developed and shown in Appendix B. The database contains many different images from outdoor pictures to computer generated vector graphics, from over exposure to under exposed images, images that simulate surveillance imagery and images with high structural details. All factors will help provide a overview of the performance of the image separation and the quality of the separation.

For the purpose of this test, several images were selected from Appendix B. The selected images used for this test varied from under to over exposure, from different acquisition methods, compressed to uncompressed and vector as well as raster images. The small subset of images from which the optimum sample size was determined are shown in subfigure (a) of Figures D.1 - D.6, in Appendix D.

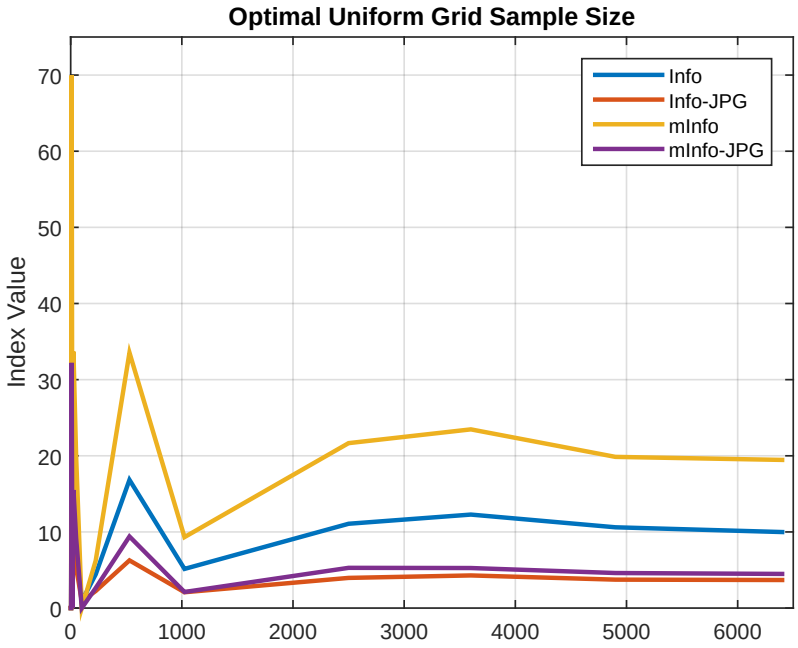


Figure 7.9: Optimal sampling size results for the uniform sampling method, for the worst case scenario of image "BGO_000[nfov]"

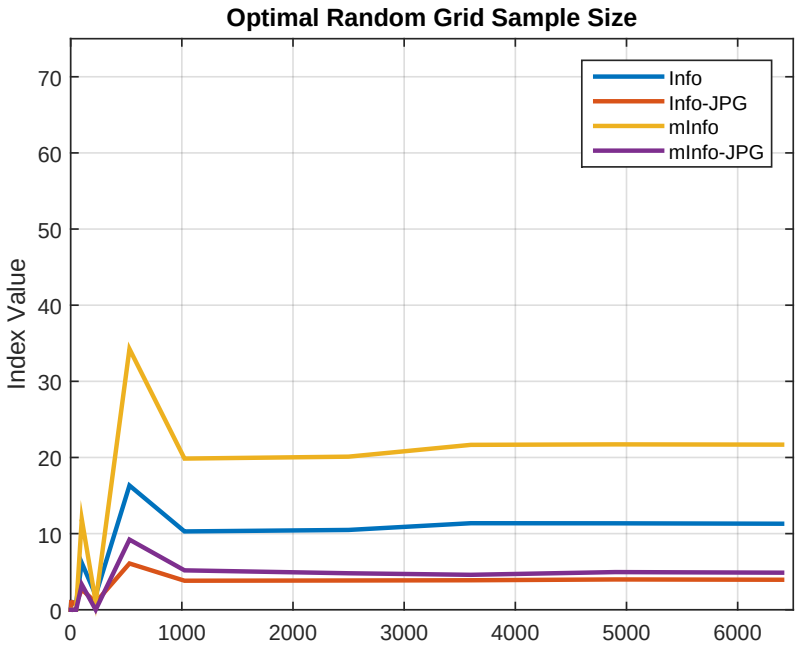


Figure 7.10: Optimal sampling size results for the random sampling method, for the worst case scenario of image "BGO_000[nfov]"

Figures 7.9 and 7.10 show the variation of the index values as the sample size is increased. The results above present the worst results from the dataset analysed. The results show that there is little significance between sampling methods and that a sampling size of 5000 patches is sufficient to provide good quality results for various different image types, with minimal performance loss. From the analysis of this study, the decision to use a suitable sample size of 5000 patches with a random sampling strategy, provides the best computational performance whilst providing reliable results.

7.5 Image Quality Assessment for Temporal Averaging Separation

The aim of this section is to analyse the quality of the temporal averaging (TA), multi-frame separation technique using several industrial standard image quality assessment (IQA) methods available, as well as the proposed image quality information metric (IQIM) technique proposed earlier.

The use of FR-IQA and NR-IQA methods will be used to give a broad scope of measurements. The simulations will be run over the entire image database containing over 30 different high resolution images (all images are above 2048×2048 pixels) from different sources and with varying image characteristics, as shown in Appendix B. The super-resolution technique will used be once again, as a benchmark method to compare our findings, and to evaluate the performance of the IQIM technique.

The selected IQM metrics are: MSE, PSNR, Pixel intensity difference, SSIM, MS-SSIM, VIF, SSEQ and the proposed IQIM technique. These methods have been outlined earlier on, in the chapter (Section 7.3.1).

These metrics have been chosen because they offer the best reliability and accuracy and correlate highly with the human vision system perception. They are publicly available through their respective online resources, and are well documented and widely used in the research community and commercially.

Identical to the process described in section 6.2, each frame is offset by a random translational and horizontal shift, before any processing has been carried out. The performance of the temporal separation technique is further analysed with respect to the registration error, to assess the quality of both the super-resolution (SR) and the temporal averaging (TA) separation techniques using different images with varying characteristics. The image quality will be assessed using the most reliable and accurate IQMs publicly available, as well as our IQIM method with varying image registration errors using our wide collection of images.

7.5.1 Image Registration Performance

The multitude of figures shown in Appendix C, presents the processed results for 2 of the simulations that were carried out. In actuality, over 30 simulations were carried out, over the entire image database (Appendix B) of high resolution images.

The results of the simulations for the entire image database of 30 images, resulted in a large amount of data to analyse. Each image was processed for 50 realisations with varying registration errors from perfect registration up to a registration error of 1 pixel with over 8 different IQMs to assess the image quality. The results have been efficiently presented such that the best and worst case results will only be presented but the results of the other 28 remaining images in the database, are well represented by them.

Although, only the data for worst and best case results are displayed, 6 of the images which represent a variety of different image characteristics have been presented in Appendix D, in order to analyse and discuss, the different image properties and how they affect the recovered image quality produced by the different techniques.

The results of the 6 selected case studies, represent a wide variety of image characteristics such as under/over exposure, high/low contrast and varying structural information contained within the images, as well as different file formats (i.e. TIFF, JPEG, PNG), acquisition methods (e.g. camera or online) and compression types. This allows us to reduce any bias that may be present in the separation process to specific image types and file formats, and also enables to assess the performance of the technique and the resulting image separation quality. The images are obtained from several different sources, some are generated from vector images, others are raster/bitmap images,

whilst others have been obtained using a modern DSLR camera (Casio EX-V7), as well as from an ordinary point and shoot camera (Nikon L27). Some of the images have also been obtained from online resources, such that the source of the image and processing they may have undergone are completely unknown. The figures in Appendix D, show the recovered $nFoV$ images using the SR and TA techniques with varying registration errors.

MSE

The Mean Squared Error (MSE) data shows that as the number of realisations increase, the MSE decreases. Initially, as the number of example images increases there is a rapid decrease in the MSE ($\approx 1-10$ realisations) after which the improvement in the MSE plateaus. Comparing the TA and SR techniques with respect to the registration error, it is clear that for registration errors of $\sigma_{err} < 0.2px$, the SR technique provides better image recovery (between $5-10$ realisations), but for a registration error of $\sigma_{err} \geq 0.2px$, the TA technique provides similar, if not better image quality (see Figure 7.11).

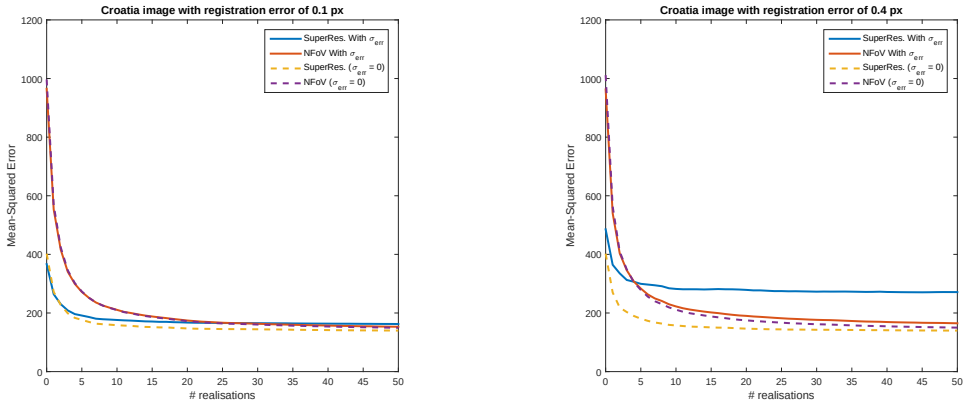


Figure 7.11: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

The vector images, "BGO_000[nfov]" and "night-cityscape" show differing results. The night-cityscape image (Appendix B, Figure(28)) is a landscape image of an arbitrary city with tall buildings and varying contrast from the side to the centre of the image. Both the $wFoV$ and $nFoV$ images produced by the super-resolution and temporal averaging techniques are identical and almost perfect reproductions of the original images according to the MSE as there is almost zero error. Furthermore, it can be seen that the SR technique does suffer some loss in image quality as the registration error increases, as shown in Figure 7.12. Visually this can be seen more prominently by comparing Figures D.4(d) and D.4(e), in Appendix D.

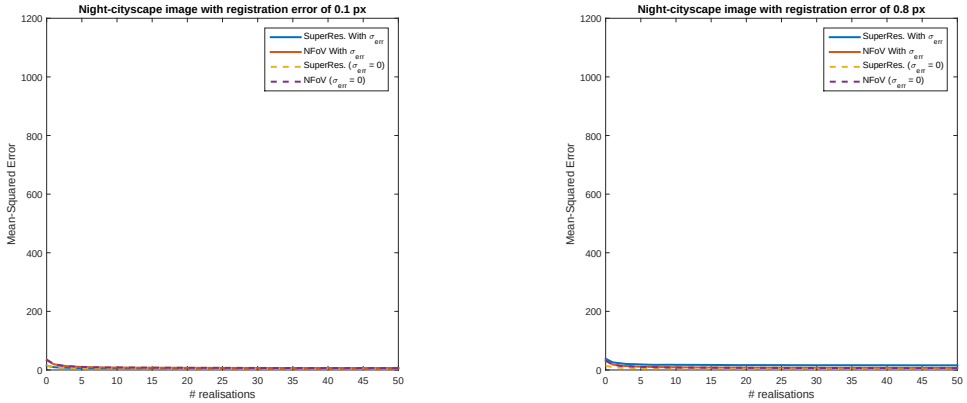


Figure 7.12: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.8px$ using the Night-cityscape image (Appendix B, Figure(28))

The "BGO_000[nfov]" (Appendix B Figure(3)) image, initially shows a large error for the $nFoV$ images, in both the SR and TA techniques. This is due to the image separation technique, which is known to have difficulty separating images when there is a large contrast difference between the centre pixel and its neighbouring pixels. This vector image was created for the specific testing of this known behaviour using Sketchup Pro© [156], to understand how well the TA technique functions. From the figures it can be seen that after $5 - 10$ realisations, the error significantly drops and after 10 realisations the results are comparable to the SR method. The two vector images also show that even with registration errors, both techniques can produce almost identical images when compared to the original $nFoV$ image according to the MSE. Visually, from Figure D.6, in Appendix D, it can be seen that the macro-blocking is present throughout with the TA technique, even though the "+" symbol has been recovered with more clarity using the TA technique (Figure D.6(d)).

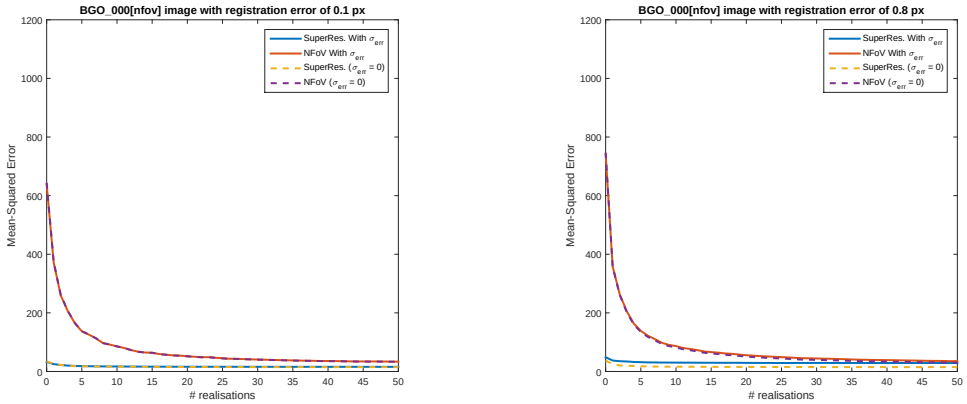


Figure 7.13: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.8px$ using the BGO_000[nfov] image (Appendix B, Figure(3))

PSNR

In general the peak signal-to-noise ratio (PSNR), shows that for $\sigma_{err} < 0.2px$, the SR technique provides better image quality, but for cases where $\sigma_{err} \geq 0.2px$, both methods provide similar results according to the PSNR. For all σ_{err}, both techniques struggle to improve the quality of the recovered images after $5 - 10$ realisations. Analysing the PSNR figures in Appendix C between $5 - 15$ realisations. It can be seen that as the registration error (σ_{err}) increases, the TA technique tends to provide a higher PSNR value and hence better image according to this IQM.

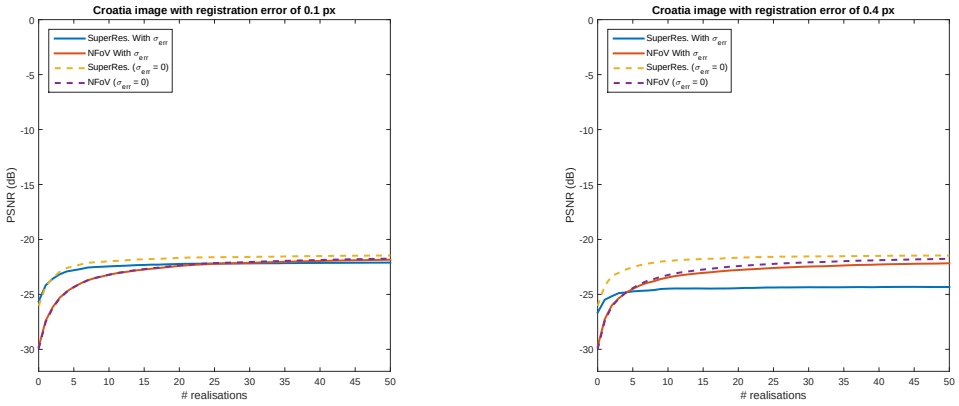


Figure 7.14: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

The vector image created in Sketchup© (i.e BGO_000[nfov]), shows that the TA technique struggled to recover the quality of the image. This was due to the inability to recover the individual frames correctly, because of the large contrast difference between the centre pixel and its neighbouring pixels. In this case, the SR technique will naturally produce better quality images because no image separation technique is used, unlike in the TA technique (see Figure 7.15). It can be seen that as the image registration error increases, the TA technique begins to approach the quality of the SR technique, as the number of realisations increase. This is because the centre of the image is offset, due to the registration errors, and the larger the error becomes the further away the centre pixel moves away from the centre, thus, allowing for better image recovery from the single frame separation and furthermore from the temporal averaging (TA).

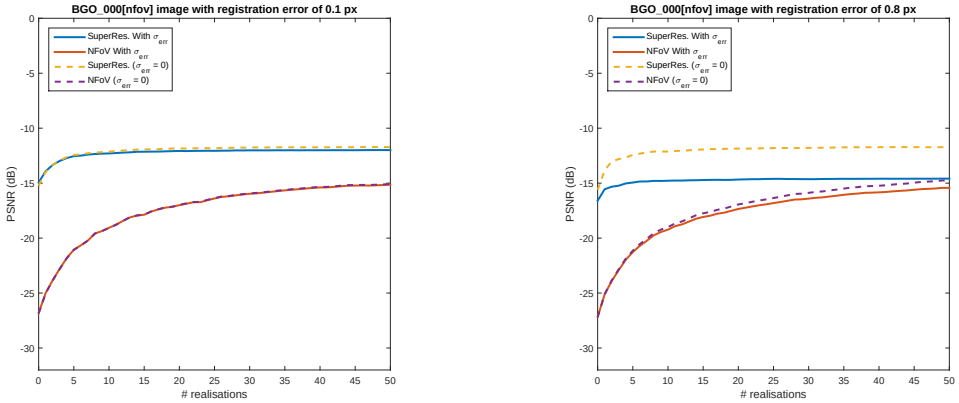


Figure 7.15: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.8px$ using the BGO_000[nfov] image (Appendix B, Figure(3))

Averaged Pixel Intensity Error

The average pixel intensity error measurements shows a similar trend as the PSNR: as the number of realisations increase the average pixel intensity error reduces. After $5 - 10$ realisations, the improvement in image quality begins to plateau. The vector image (night-cityscape) produces the best images, with the lowest errors across all registration error (σ_{err}) values. For $\sigma_{err} < 0.2px$, the SR technique provides the best quality images but, as the number of realisations increases, the TA technique is able to produce similar quality results. When $\sigma_{err} \geq 0.2px$, the TA technique begins to produce equal if not better results with respect to this IQM. As the σ_{err}, increases the TA technique provides better image quality and shows to be more robust to registration errors, compared to the SR technique and when compared to the perfect registration case. This behaviour is particularly prevalent for results where $\sigma_{err} > 0.4px$. This can be seen, visually by looking at Figure D.2 in D.

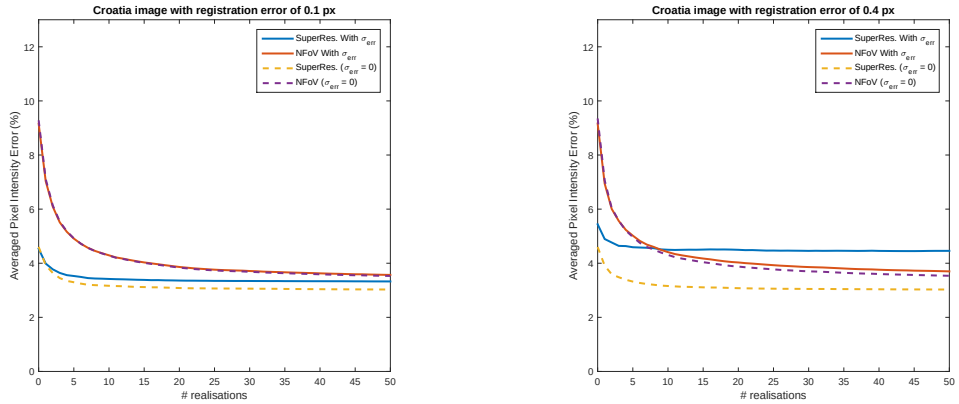


Figure 7.16: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

SSIM

The structural similarity index (SSIM) results follow the same trends as the PSNR and averaged pixel intensity error IQMs. The quality of the images improve as the realisations increase. Little improvement in image quality is achieved after $5-10$ realisations. The vector image (night-cityscape) provides the best image quality recovery. The TA method provides better image quality for $\sigma_{err} \geq 0.2px$ after just 5 realisations. It also shows the TA technique is more robust to registration errors because there is a negligible difference ($\approx 1\%$) between the results at all σ_{err} values compared with perfect registration results. In this situation, the SR technique struggles to perform as well. The CAD generated vector image is once again the weakest of the results. Comparing the TA technique with registration errors to the perfect registration case, illustrates that the TA technique is robust to registration errors, but has difficulty performing the single frame separation due to the large contrast difference of the centre pixel with its surrounding pixels.

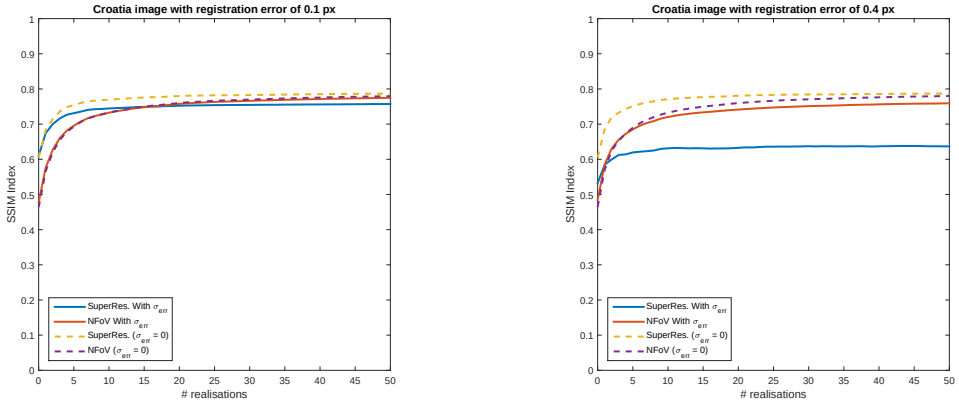


Figure 7.17: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

MS-SSIM

The multi-scale structural similarity (MS-SSIM) index, is an extension of the SSIM IQM as described in Section 7.3. The results obtained from the MS-SSIM index describes the same findings and trends, as found with the SSIM index.

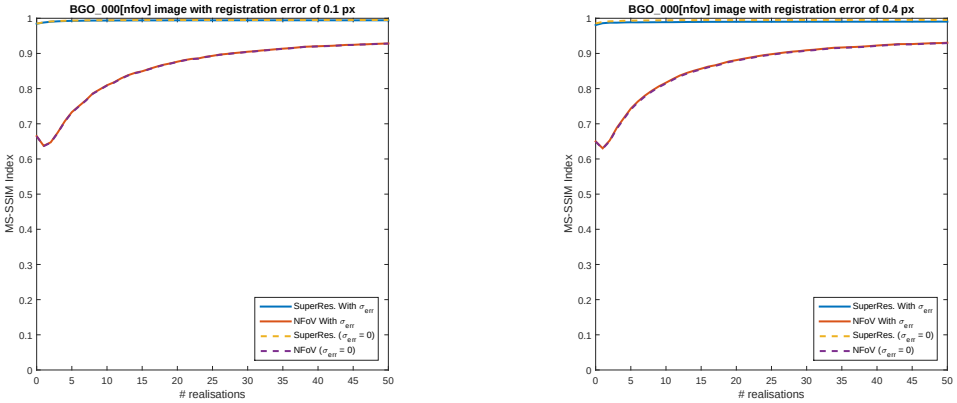


Figure 7.18: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the BGO_000[nfov] image (Appendix B, Figure(3))

Figure 7.18 shows that in this case, with the vector image, BGO_000[nfov] the SR technique outperforms the TA technique, as already described, in the SSIM section.

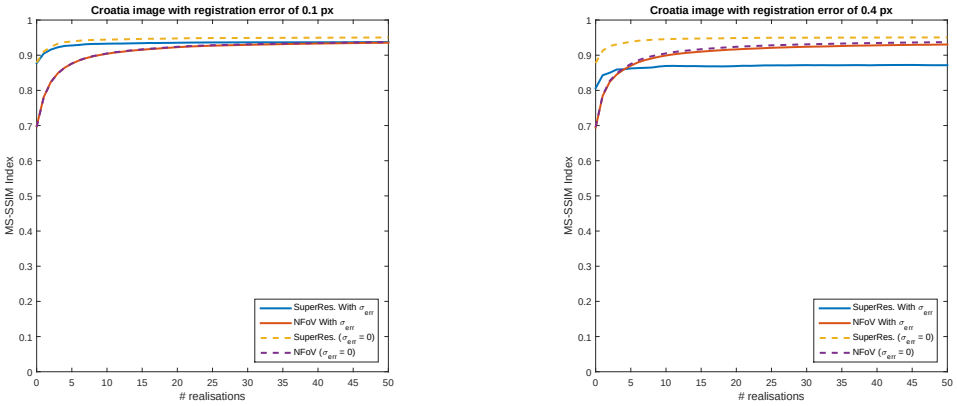


Figure 7.19: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

VIF

The visual information fidelity (VIF) IQM, shows the image quality improves as the number of realisations increases. After $5 - 10$ realisations, the improvement in image quality plateaus and little improvement in quality is gained, but at the expense of computational costs. For $\sigma_{err} < 0.2px$, both the SR and TA techniques provide very similar results, in terms of the quality metric. At $\sigma_{err} \geq 0.2px$, the SR technique begins to suffer in comparison to the TA technique; it can be seen that as the registration error increases the SR image quality degrades providing lower quality images. The TA technique provides better image quality at registration errors larger than $0.2px$, but also for lower registration errors, after $5 - 10$ realisations, the TA technique can provide similar quality images. The TA technique also provides better quality images at registration values $\geq 0.2px$, very close to that of the images with perfect registration,

i.e $\sigma_{err} = 0px$, showing that the TA technique is more robust to changes in registration error and in most cases able to provide images of quality similar to the SR technique.

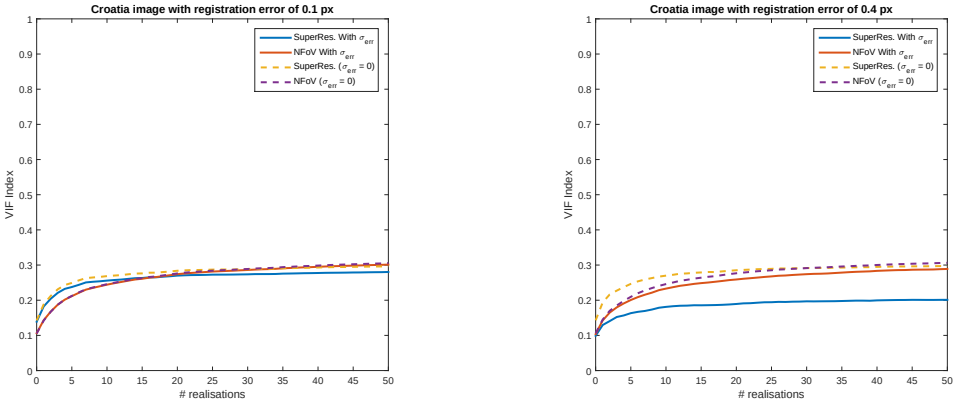


Figure 7.20: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

SSEQ

The NR-IQA spatial-spectral entropy-based quality (SSEQ) IQM, is a training based model. For the SSEQ index, 0 represents the worst quality and 100 represents the best. The results for SSEQ show that there is slight fluctuation, in the image quality between $1 - 5$ realisations, after which there is little to no improvement for both techniques, at all registration error values. The SSEQ has shown to be insensitive to registration errors, this can be seen, because the perfect registration cases are almost identical to the measurements with registration errors, but when visually assessing the recovered images in Appendix D, it can be seen that there are visual differences. This may be due to the SSEQ metric being a training based method, and specific training for registration errors were not considered. The TA technique provides similar image quality in comparison to the SR technique at all registration error values, according to the SSEQ metric. The perfect registration cases show that there is very little difference between the perfect registration case and the case with registration errors. The vector images show the same findings, but the results for the SR technique is much better in comparison to the TA technique because of the problem with the single-frame separation. The TA technique is able to generate images of similar quality to those with perfect registration, illustrating its robustness and stability as a separation technique, whereas the SR technique does suffer some loss in quality with the introduction of registration errors.

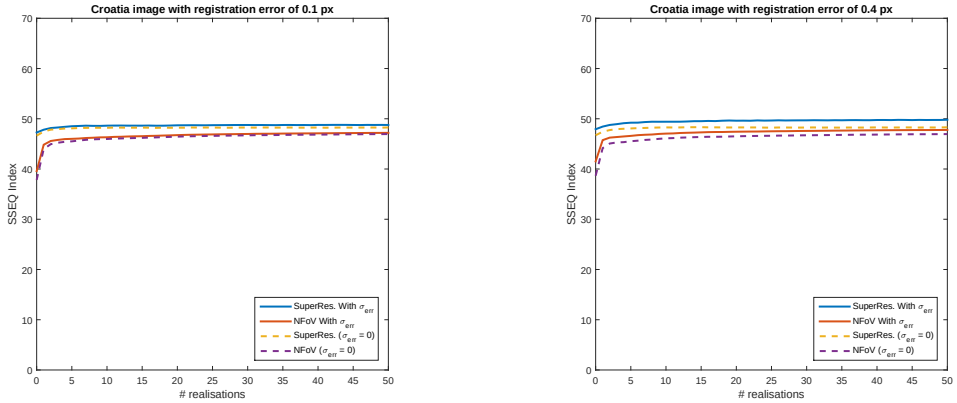


Figure 7.21: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

IQIM

The IQIM metric shows both techniques, with and without registration errors, initially start at a high value. As the number of realisations increase, the IQIM begins to reduce and eventually plateaus. The results show that after $5-10$ realisations there is no large improvement in the quality metric as presented by the figures in Appendix D.

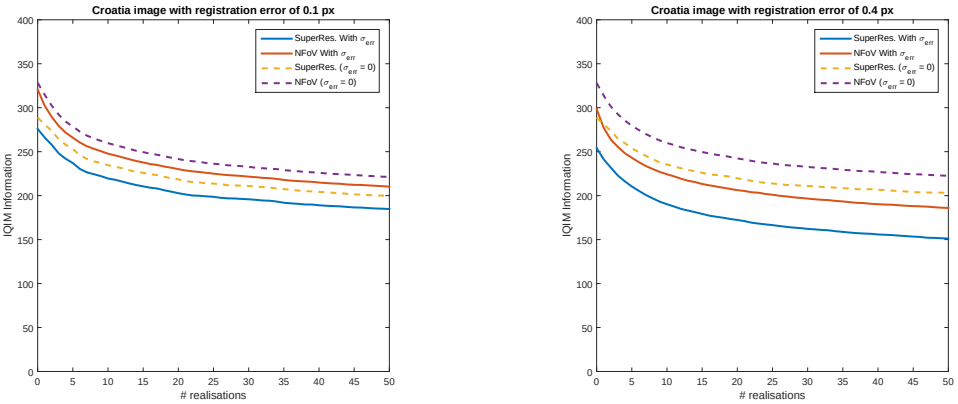


Figure 7.22: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

This shows that the quality of the recovered images have stabilised and further realisations do not substantially improve the quality of the images produced. At this point the trade-off between image quality and computational costs must be considered, depending on the application. In the case with perfect registration, it can be seen that the trend is identical for both the SR and TA techniques. The TA technique shows to be marginally better than the SR technique as can be seen by the figures in Appendix C and in figure 7.22. This shows that more information is present in the image produced by the TA technique compared to the SR technique. As the registration error increases, both the SR and TA techniques begin to suffer with loss of information in comparison

to the perfect registration cases. The figures show the SR technique tends to degrade much quicker, as the registration error increases, whereas the TA technique is able to recapture more information albeit less than the perfect registration case.

The self created CAD based vector image "BGO_000[nfov]" which is known struggle with the basic image separation technique, shows that the SR technique provides the best results throughout the simulations. The amount of information does not change as the registration error increases with respect to the SR technique. The TA technique shows that initially the amount of information increases, this could be due to the poor initial single frame separation but as the realisations increase the TA technique plateaus to is initial value. It is clear that in this case the SR technique provides far better images, due to the poor separation of the single frame technique.

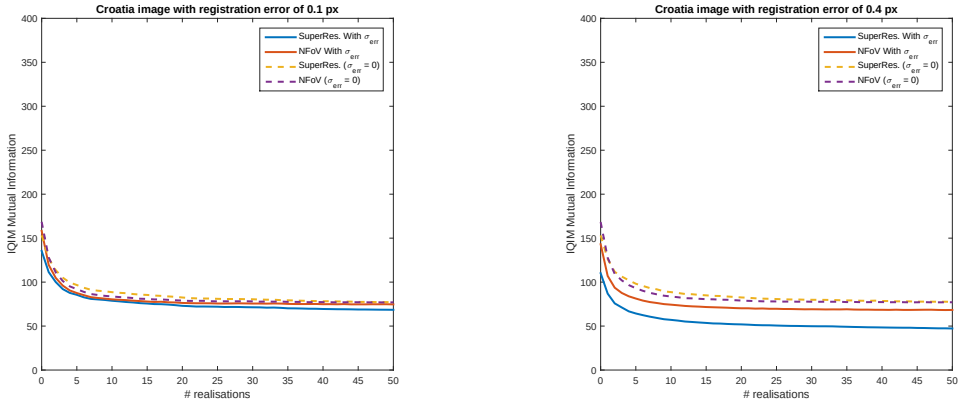


Figure 7.23: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

The IQIM shows that with perfect registration the TA and SR techniques provide similar amounts of mutual information, in fact they are almost identical. Initially within the first $5 - 10$ realisations there is a reduction in mutual information, after which the amount of mutual information plateaus. As the registration error increases it can be seen the the mutual information begins to gradually reduce further; this is more apparent with the SR technique. The trend stays the same with both SR and TA, but the SR suffers greater mutual information loss as the registration error increases. The mutual information shows that the TA technique is able to stay much closer to the perfect registration results, showing the technique's robustness to registration errors. The mutual information figures show poor separation using the TA procedure with the self-created vector image. The figures show that after many iterations (≈ 30) it is able to reach similar amounts of mutual information present as compared to the SR produced images. The IQIM metric and the mutual information do not correlate highly with the HVS. In reality, it merely shows the amount of information and mutual information present in the images produced by the differing techniques. Once the index

stabilises, it represents the best possible separation of images available. Understanding this limitation, a JPEG version of the index was produced as the JPEG matrix emphasises the frequency ranges that are most important with regards to the HVS, and will provide better estimations tailored to the HVS.

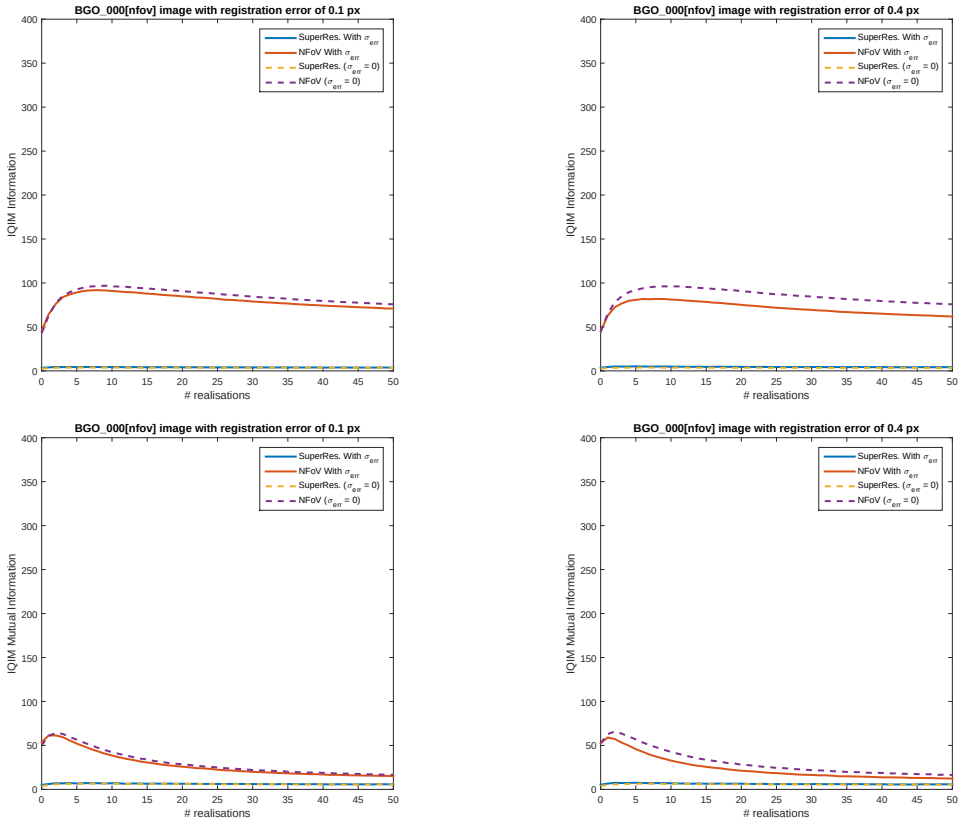


Figure 7.24: Comparison of the TA and SR techniques for $\sigma_{err} = 0.1px$ and $\sigma_{err} = 0.4px$ using the Croatia image (Appendix B, Figure(16))

IQIM JPEG

The JPEG alternative method of the IQIM index shows the same trends as discussed in the IQIM section, including the results of the self created vector image showing the poor separation using the TA technique. The magnitude of estimated information that can be perceived by the HVS is much less than the total estimated information present, as shown by the IQIM index and Figure 7.25 showing the amount of redundant information that the HVS cannot process. The same findings can be said for the mutual information as compared to the JPEG version of mutual information.

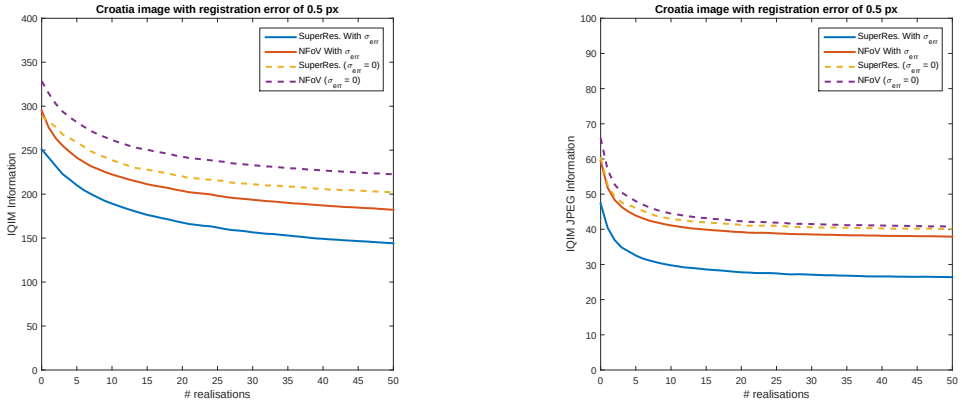


Figure 7.25: Comparison of the TA and SR techniques for $\sigma_{err} = 0.5px$ using the Croatia image (Appendix B, Figure(3))

7.5.2 Summary of Results

In summary, in most cases, except for the self-created vector image, it can be seen that after $5-10$ realisations the recovery of the image quality plateaus as we begin to reach the optimal quality possible, after which point, the computational costs will begin to increase and outweigh the improvement in quality. Another point to consider is the scenario with a live feed, where the computational time will be important to be able to generate as near real-time footage as possible.

Visually assessing the recovered images, presented in Figures D.1-D.6 in Appendix D, not only reaffirms the results presented in Chapter 6, but also shows that with registration errors due to frame-by-frame motion the temporal averaging technique can produce images with quality similar to the super-resolution technique, if not better, when dealing with large (i.e. $\sigma_{err} \geq 0.2px$) registration errors. From [20], with the optimal parameters, SURF is able to reach a registration accuracy of $0.2px$ and when dealing with 3D reconstruction registration, it can be greater than $1px$ (as illustrated in Table 2 in [20]). Comparing the results of the temporal separation against the super-resolution technique at registration errors above $0.2px$, the temporal separation technique has similar, if not better, reliability with respect to image quality between $5-10$ realisations with all images, including other vector images as shown in D.4(a), but with the exception of the self created vector image Figure D.6(a). The results also show that certain vector images tend to be difficult to separate. The self created vector image, showed that the co-boresighted requirement is a key factor to the performance of initial single-frame separation. The vector image, contains a "+" in the centre pixels, and due to the intra-pixel boundary (pixel selection) issue that the initial single-frame separation technique suffers from, and the high frequency content of the edges generated by the vector images, the technique suffers from poor separation, added with the temporal separation process, which attempts to use the motion changes between

frames, in order to improve the separation. In this particular case the errors accumulate over the averaging process, but this is not with all vector images as can be seen by the vector image D.4 and others which were tested.

7.6 Moving Average Temporal Separation

The temporal separation technique, and the results from the previous section have shown the technique to be robust to registration errors, and with between $5 - 10$ realisations the technique can provide images with similar, if not better, quality then the super-resolution technique.

The temporal separation technique was designed to facilitate fixed cameras which exhibit small variations in motion due to external factors, such as vibrations or weather. In this section, the findings from the previous section are used to extend the separation technique to work with non-fixed cameras. To enable this, a moving average technique is utilised. From the previous findings, it was shown that between $5 - 10$ realisations were needed to provide images with sufficient quality without excessive computational costs.

Knowing that the optimal number of frames to use is between $5 - 10$, the moving average technique takes the most current frame, and registers a number of the previous frames to the current frame and generates the average of them. The advantage of this method is that it allows the camera to exhibit motions such as pan, tilt and zoom and using image registration allows the recovery of any information that was present in the previous frames to produce the best available image, with minimal computational cost.

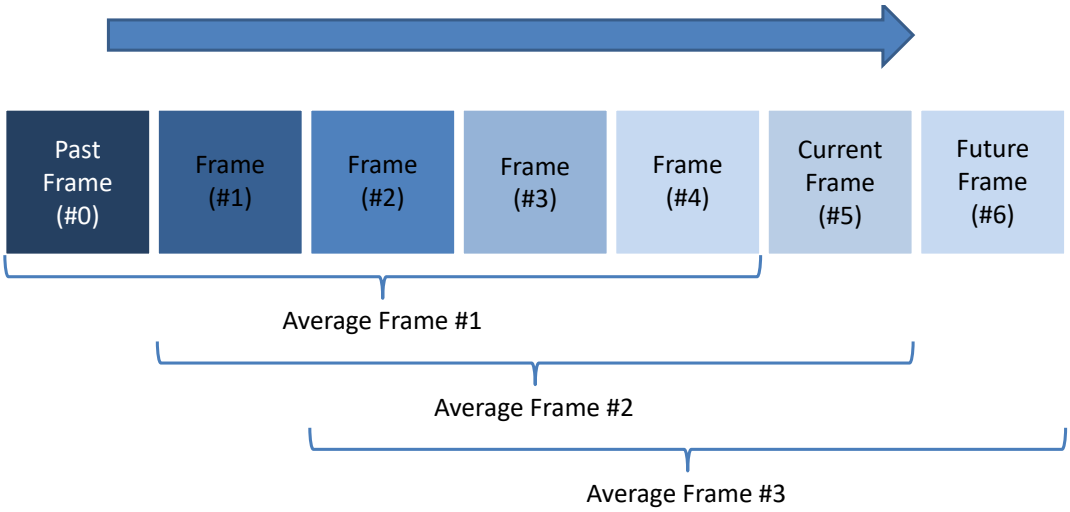


Figure 7.26: Illustration of moving average (MA) technique for video data

Temporal separation is limited to fixed cameras, which exhibit jitter. The moving average technique recognises that the benefit of averaging the previous frames with the

current frame, in order to help alleviate distortion and recover structural information. From the temporal separation we can see that significant improvement in image quality can be recovered within 5 frames when using an averaging technique, hence a 5 frame moving average was implemented.

Some results have been simulated using the Croatia image used previously in this chapter and in Chapter 6. The image was used to create a zooming effect, as if to simulate a missile system, that had 'locked on' to its target and was travelling towards it, but there was no simulated "jitter", only a smooth transition, which resulted in the prominent macro-blocking artefacts to still be present after the averaging process due to the intra-pixel boundary issues from the single-frame separation. The 5-frame and 10-frame moving average technique was applied to the recovered $wFoV$ and $nFoV$ images. The results of some of the selected frames are shown in Figure 7.27 and 7.28.

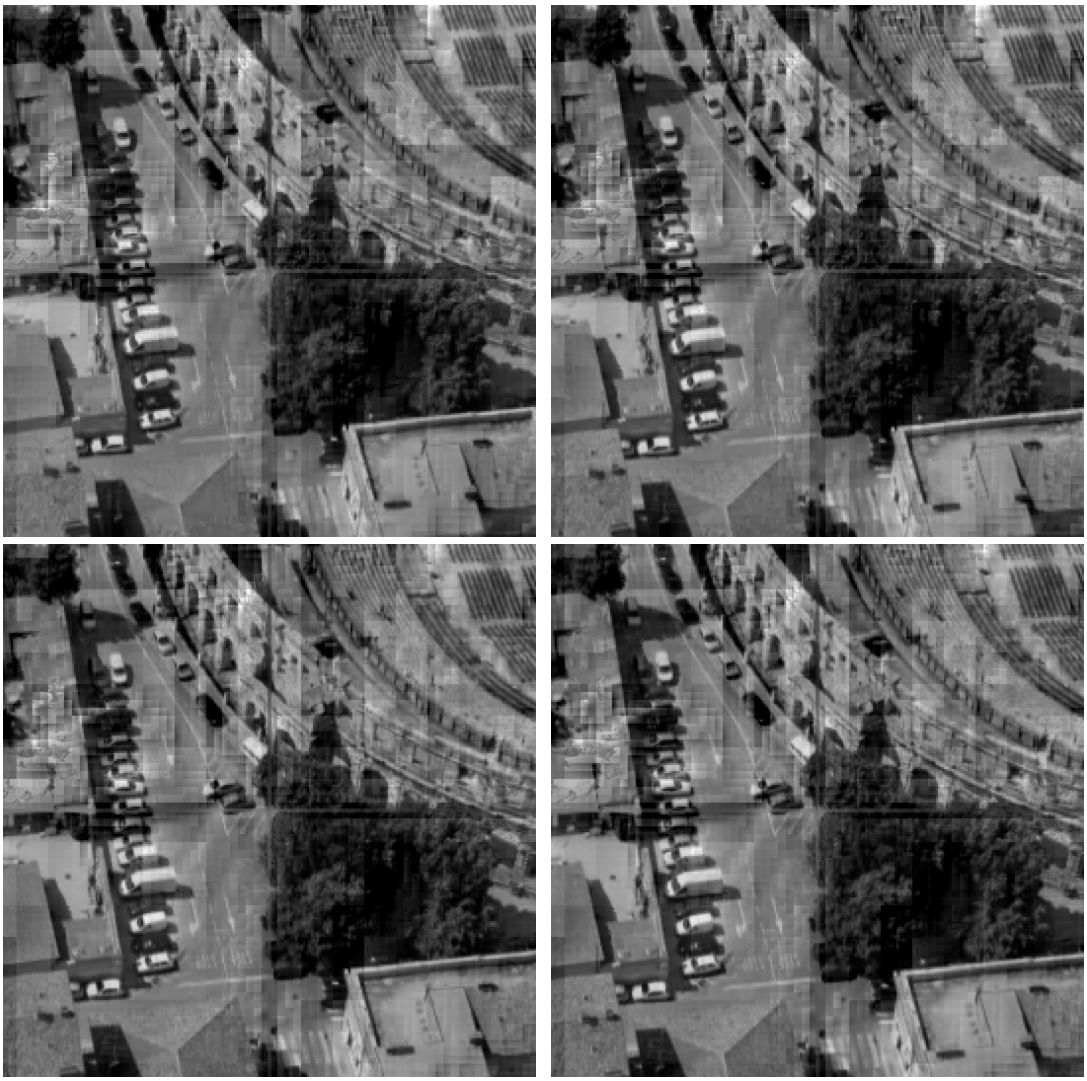


Figure 7.27: 5-frame moving average separation technique of recovered $nFoV$, using simulated camera zooming motion without "jitter". [Top Left] - frame 1, [Top Right] - frame 14, [Bottom Left] - frame 34, [Bottom Right] - frame 54

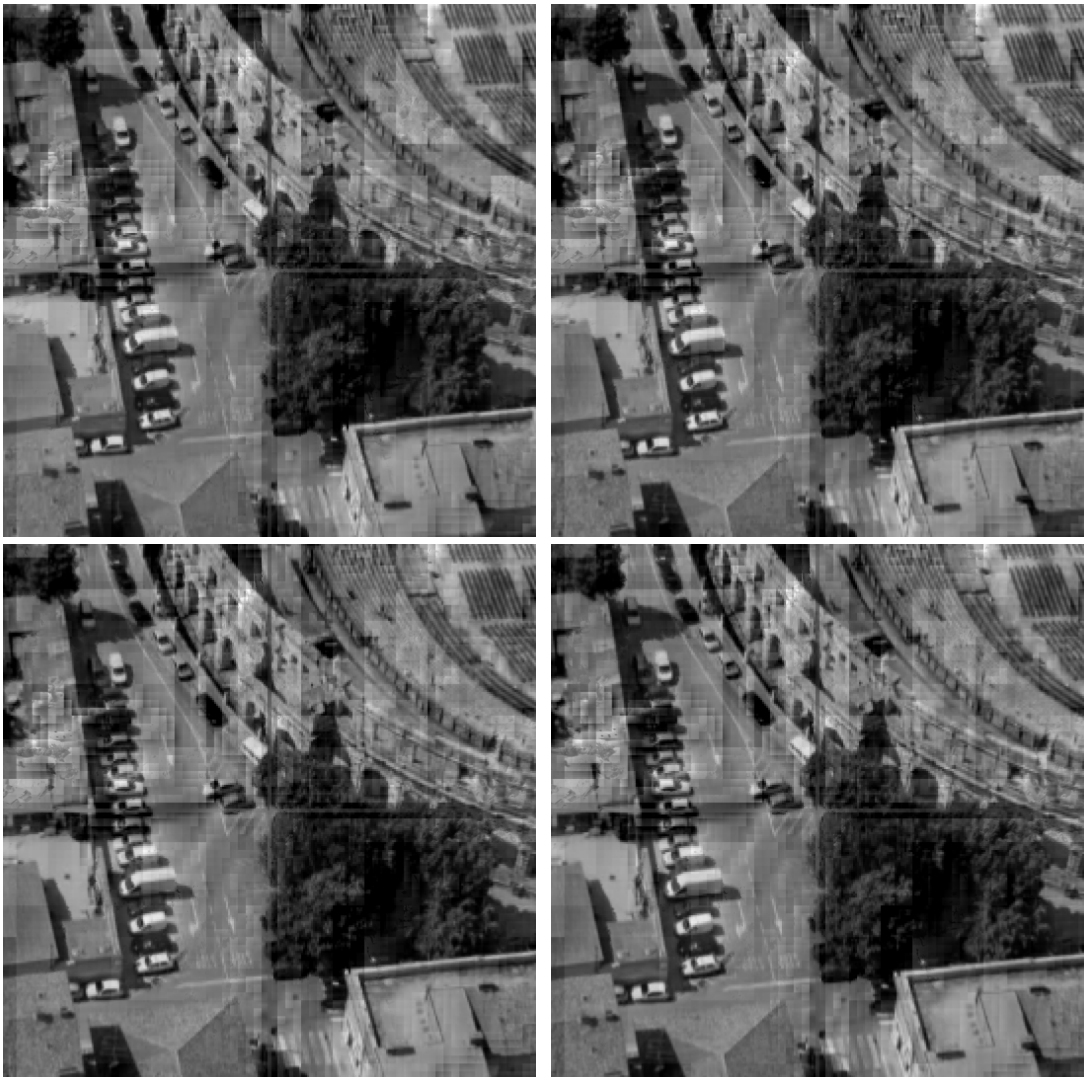


Figure 7.28: 10-frame moving average separation technique of recovered $nFoV$, using simulated camera zooming motion without "jitter". [Top Left] - frame 1, [Top Right] - frame 14, [Bottom Left] - frame 34, [Bottom Right] - frame 54

Since the publication of [179] of the single frame separation technique, there has been some interest in the dual field-of-view imaging system from Prof. Andy Harvey, School of Physics and Astronomy at the University of Glasgow [246,290] and Dr. James Babbington from Qioptiq Ltd., a commercial photonics company, interested in designing the optics required for such as dual field-of-view imager. See Figure 7.29, which presents one of the proposed optical designs to superimpose the $wFoV$ and $nFoV$ coboresighted images together. The optical design in Figure 7.29 shows two separate lenses, one for the $wFoV$ and another for the $nFoV$, but another design is currently being developed to only have a single lens, with two optical paths for the $wFoV$ and $nFoV$ imagery which will be superimposed onto a single imager.

7.7 Chapter Summary

In summary, this chapter provided a broad case study of the Temporal averaging (TA) initially proposed in Chapter 6. The TA techniques facilitates fixed cameras which exhibit small desired/undesired camera motions such as vibrations, which produces "jitter". The broad case study was carried out on over 30 high resolution ($2048px \times 2048px$) images. The results showed that for most images, the TA techniques is robust to registration errors, within $5-10$ realisations and can perform to a similar standard, if not better, compared to the super-resolution technique. Current image registration techniques have been shown to be accurate to $\sigma_{err} \geq 0.2px$, and in these cases TA has shown to provide good, if not better, image quality within $5-10$ realisations. The TA technique suffers poor separation performance with vector images, which contain high contrast objects, such as points, lines or curves near the centre pixels and its surrounding region. In these cases the single-frame separation technique suffers from intra-pixel boundary issues (as described in Chapter 4), which manifests itself as macro-blocking artefacts.

The quality of the recovered images were quantitatively analysed using a range of available image quality metrics (IQMs), which were also reviewed in this chapter. A new blind IQM was also proposed in this chapter called "image quality information metric" or IQIM, which determines the amount of information available in an image using the DCT domain. The metric was shown to not correlate highly with the HVS, so another variant was proposed which utilised the JPEG matrix, which keeps the frequency information content that the HVS can interpret and removes the information that cannot be processed by the HVS. However, the results were the same but showed much less available information content for each image.

Using the information provided by the case study shows that, in most cases, after $5-10$ realisations the TA technique provides sufficient image quality without excessive computational costs being required. A moving average technique was subsequently implemented taking the current frame and its previous captured frames and using image registration to create the best image available. Some results have been illustrated in Figure 7.27 using a 5-frame moving average technique. The moving average technique will allow the separation technique to facilitate motions that camera systems on-board an aircraft may undergo, such as zooming, tilting, panning etc. whilst also maximising the information content and improving the separation process through desired/undesired camera motion, i.e. "jitter", with the proposed $dFoV$ imaging system.

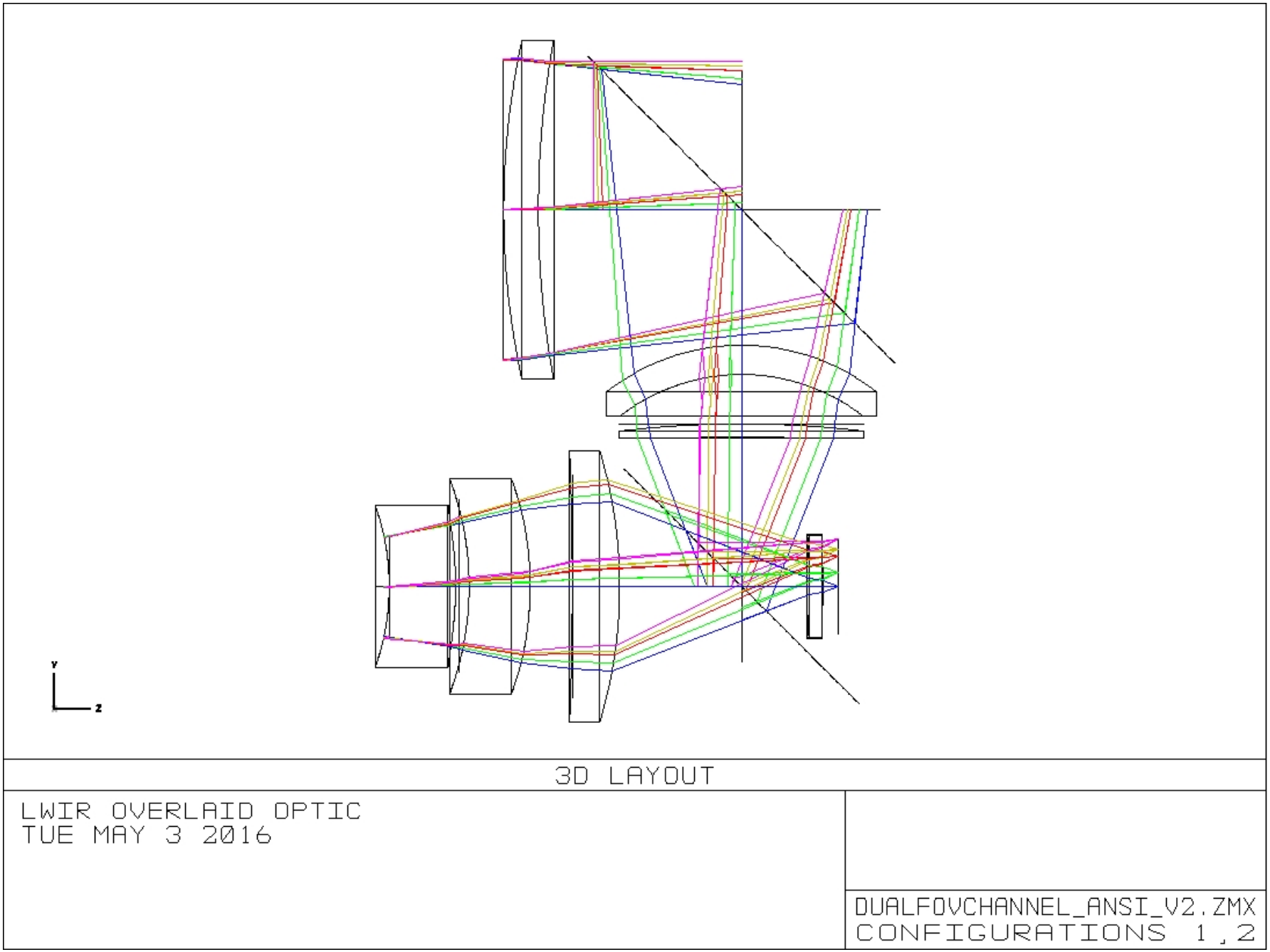


Figure 7.29: Hardware design for the potential lens system for the proposed *dFoV* imaging system

Chapter 8

Summary & Conclusions

8.1 Summary

This thesis focused on a novel dual field-of-view ($dFoV$) imaging sensor/seeker for application to missile systems and unmanned aerial vehicles (UAVs). The proposed imaging sensor, simultaneously captures both the narrow field-of-view ($nFoV$) and the wide field-of-view ($wFoV$) imagery, of the same scene, co-boresighted and superimposed, on to a single superimposed image frame. A geometric separation technique was proposed to recover the mixed images, into its constituent $wFoV$ and $nFoV$ imagery, using only the geometric relationship between the two fields-of-views. The quality of the recovered images was assessed using current image quality analysis techniques. Additionally, the work carried out on image quality analysis, led to the development of a new image quality metric. The proposed metric's performance was analysed against current image quality metrics. The geometric separation technique was also extended to work with video data, for fixed cameras that exhibit "jitter" and later to operate with moving cameras.

The advantages of a $dFoV$ imaging system allows to transmit a single mixed image to the remote operator, once received the constituent $wFoV$ and $nFoV$ imagery can be recovered. This can provide the remote operator, to provide surveillance capability and target detection using the $wFoV$ imagery, and at the same time; identify and track potential targets using the $nFoV$ imagery. An additional advantage of the superimposed imagery, is the ability to transmit two video streams at a fraction of the bandwidth required to transmit two individual video feeds (refer to Appendix A).

Chapter 2 outlines the theory required to develop a 6 degrees of freedom simulation of a missile airframe. The missile model was created to generate realistic trajectories in order to understand the motions its onboard imaging system/seeker would undergo, in a safe digital environment. The performance and stability of the model was also analysed.

Chapter 3 extends the previous chapter to evaluate the current-state-of-the-art seekers and guidance systems. The implementation of the laser beam rider (LBR) system

was chosen, for application with the $dFoV$ imaging system, as it is a reliable and highly resistant to counter-measures, guidance system. The laser beam riding system continues to grow in popularity, for current and future military systems as shown by the FASGW(L) projects, funded in partnership with the UK MoD.

Chapter 4 introduces the single-frame geometric separation technique. The co-boresighted, superimposed imagery and the geometric relation between the two field-of-views are used to recover the $wFoV$ and $nFoV$ images produced by the $dFoV$ imaging system. The separation technique has been evaluated against the benchmark Lena test image and shows good recovery, but macro-blocking artefacts were found to be present due to intra-pixel boundary problems.

Chapter 5 reviews the range of image registration techniques currently available from area-based to feature-based methods. The review was carried out in order to select the best solution for application to the $dFoV$ imaging system, and to handle video data onboard a missile system, which will transmit the imagery to the remote operator. The speeded-up robust features (SURF) detector and descriptor was used in conjunction with the random sample consensus (RANSAC) algorithm to match and estimate the geometric transform between sequential frames. The SURF technique has been shown to provide robust feature selection and detection and provides good computational speed and efficiency, compared to other feature-based methods. For real-time or near-time situations, computational speed is important, SURF provides accuracy similar to SIFT with reduced computational time. Image registration was implemented and a brief analysis of the potential bandwidth savings was reviewed which shows at least 40% reduction, in our test, with the best case showing 60% bandwidth savings over streaming two individual video streams (in Appendix A).

Chapter 6 presents the multi-frame temporal-averaging geometric separation technique. The temporal averaging technique extends the single-frame geometric separation technique to function over fixed cameras, that suffer from motion due to external influences e.g. weather. The temporal averaging technique's separation performance and image quality was analysed using the popular structural similarity (SSIM) index. The performance of the temporal averaging technique has been compared against the super-resolution technique as a benchmark to assess the image quality. The results show that the temporal averaging technique produces images of similar quality to the benchmark super-resolution technique, but offers better robustness and separation performance, when exposed to registration errors, where the temporal averaging technique begins to outperform the super-resolution technique around $0.1px$ to $0.2px$ registration error.

Chapter 7 follows from the previous chapter. The analysis of the temporal averaging technique is broadened to assess the image quality of the method against a collection of over 30 high resolution ($2048px \times 2048px$) images and the sensitivity to registration errors has also been assessed. The super-resolution technique was used once again

as a benchmark. The introduction of a new image quality metric (IQM), the image quality information metric (IQIM), was presented and analysed along with several other IQMs, which were reviewed earlier in the chapter. Overall, the results show that for perfect registration the super-resolution technique outperforms the temporal averaging method within the first $1-5$ frames, after which the temporal averaging method produces images of similar quality, and in some cases better quality, as the number of frames increase. As the registration error increases, most noticeably, beyond $0.2px$ error, we can see that the temporal averaging technique begins to produce better quality images after $5-10$ frames and shows far better robustness to registration errors compared to the super-resolution technique. Previous studies have shown that at best, the SURF method can be accurate to $0.2px$ error [20]. All of the chosen image quality assessment techniques support these findings. The results also show that the temporal averaging technique is considerably more robust to registration errors compared to the super-resolution technique. This is because the super-resolution technique extracts new information from small changes in the frames, but as the registration error increases, this become increasing difficult. Following on from this, the temporal averaging method was extended to work for moving cameras. The earlier findings show that using SURF which has an accuracy up to $0.2px$, and that roughly $5-10$ frames are needed to produce images of similar quality, in comparison to the benchmark super-resolution technique, the moving average separation technique is developed to accommodate for moving cameras (i.e. panning, tilting, zooming etc.).

This thesis has focused on a novel dual field-of-view imaging system which could lead to simultaneous, $wFoV$, surveillance and detection and $nFoV$, target identification and tracking applications. The thesis has demonstrated that it is possible to separate a superimposed image, knowing no information of the scene. It has been demonstrated, that using only the geometric relationship between the field-of-views, of the superimposed image, the respective $nFoV$ and $wFoV$ images can be recovered and used for operational purposes. The quality of the recovered images were analysed, and a new image quality metric was introduced, and its performance was analysed against current image quality metrics. It has been briefly demonstrated that there are also potential bandwidth savings, of between $40-60\%$ (see Appendix A) but this was not the main scope of the thesis and only a small study conducted. It has been shown that the initial single frame separation technique can be extended to function with multi-frame (video) situations with fixed cameras and also moving cameras. Some of the work in this thesis, has also been published in the 2014 and 2016 IEEE Information Fusion (FUSION) International conferences [179, 180].

The thesis also demonstrates that a $dFoV$ imaging system can been designed. A detailed engineering drawing of the proposed $dFoV$ imager has been created by Prof. Andrew Harvey of The University of Glasgow, in conjunction with Dr. James Babington

of Qioptiq Ltd. (see Figure 7.29). The design proposes a dual lens configuration, but preliminary designs for a single lens configuration are being actively developed.

8.2 Limitations

A limitation to the overall imaging system is the requirement of $nFoV$ and $wFoV$ optical paths to be co-boresighted when the superimposed image is created. The single-frame geometric separation technique is dependent on this assumption, as the location of the corresponding pixels and intensity values (intra-pixel boundary issue) are needed for successful separation. Moreover, it has been shown that a co-boresighted lens system can be developed, and the geometric relationship between the field-of-views can be determined and calibrated during the manufacturing stage. The single-frame geometric separation technique has been shown to encounter difficulty with images containing high frequency content (i.e. sharp edges, lines etc.) because of the pixel selection issue, and high contrast changes. However, this is most prevalent with vector images, especially when the vector image contains very high frequency information at the centre of the image, as discussed in Chapter 6 and 7. Vector images of landscapes and buildings have been successfully separated, as shown in Chapter 6. The real-world operational use of the $dFoV$ imaging system will mean that it would be unlikely to encounter such a situation, as most situations will deal with natural scenes.

Another limitation was the use of simulated imagery and video data. As the hardware capability has not yet been developed, although plans to design a lens system have begun. This has limited the results to imagery captured from single optic, camera systems and the superimposed imagery was simulated, by taking a high resolution $wFoV$, from different acquisition methods i.e. DSLR, point-and-shoot camera, online etc. and creating a simulated $nFoV$ image.

8.3 Recommendations For Future Work

The development of the $dFoV$ imaging system in conjunction with Prof. Andy Harvey of University of Glasgow and Dr. James Babbington of Qioptiq Ltd., would allow future testing of the separation technique's performance with real imagery. The designed imaging system would require the separation technique proposed in this thesis, to recover the constituent $wFoV$ and $nFoV$ images. The integration of the software (separation technique) and hardware (imaging system) would advance the research, and enable us to work in collaboration, towards a complete imaging system for real-world testing and implementation.

The single-frame separation technique has been known to create macro-blocking artefacts, mainly due to high frequency content (i.e. lines and edges), especially when there is a high contrast change because of the intra-pixel boundary and selection of the

wrong pixel intensity value. Research into the minimisation or removal of this issue, could lead to improved image separation. Suggestions to modify the approach would be to use a bottom-up approach, instead of working from the outer pixel towards the centre pixel to form the initial closed set of equations, by working in the reverse (from the centre pixel to the outer pixel) this could reduce the intra-pixel boundary issue, as the centre pixel will always be the same and hence reduce macro-blocking artefacts.

Another recommendation would be to consider using a form of computational imaging/machine learning technique, to reduce the intra-pixel boundary issue. By knowing the relationship between the two field-of-view, and the resolution of the superimposed image captured, the pixel mappings required to calculate the pixel intensity values for the $nFoV$ and $wFoV$ can be pre-determined, which would reduce the time needed to generate the closed set of equations, required to solve for the pixel intensity values at each individual pixel and reduce the pixel boundary issues.

The performance of the image separation technique must also be taken into consideration. For applications onboard aircraft, in particular missile systems and UAVs, where time sensitivity plays an important role in the decision making process, real-time, or as close as real-time solutions are necessary. The separation algorithms were coded using MATLAB© which is considered a proprietary software language, suited for development and debugging but for operational use, it would be beneficial to convert the code using a lower-level language such as C, which provides better access to computational resources and much improved performance.

Another performance consideration would be to use the computational power of the graphics processing unit (GPU), which can be used for high volume, light computational calculations, due to its access to a large number of GPU threads. This would considerably improve the image separation performance. Another method would be to develop an FPGA based design which would be better suited to low power missiles or micro-UAVs.

The continuation of the future work described, could lead to a complete $dFoV$ imaging system; hardware and software seamlessly integrated into a fully functioning system, which could have the potential for real-life testing and installation on-board an aircraft, such as a missile system or UAV.

Appendix A

Bandwidth Saving Case Study

A.1 Introduction

The proposed imaging system/seeker has the ability to superimpose both the $wFoV$ and $nFoV$ images of the same scene and transmit the single superimposed image frame to the remote terminal operator. As the two frames ($wFoV$ & $nFoV$) are superimposed before transmission there is an inherent saving of transmission bandwidth usage. In this section, the bandwidth saving will be analysed to determine the reduction in bandwidth consumption.

For this study, several outdoor videos were captured using a Nikon Coolpix L27 16.1 megapixel digital camera with a CCD imaging sensor [198]. The video selected for this study, was taken in a park, with the camera panning across at an unknown and variable speed. Initially, the video was captured by two Nikon Coolpix L27 cameras using a dual camera mount. One camera records the $wFoV$ and the second camera the $nFoV$. Unfortunately it was found to be extremely difficult to calibrate both cameras to be co-boresighted. Therefore, in order to meet the conditions required for the geometric separation technique to function correctly, the $wFoV$ video was taken and the central portion of the video was extracted and scaled to create an augmented $nFoV$ video. This limits the high frequency content available from the $nFoV$ video data, but for this study the high frequency information is not essential to this study.

Digital images are numerical representations of a two or three dimensional image/video. There are two main forms of digital images: **raster** and **vector** images. Raster images are images of fixed resolution i.e. a set number of pixels. The digital raster image is represented as an array with a fixed number of rows and columns. Each value in the array corresponds to a pixel where the value is representative of the brightness of the pixel of a given colour. Vector images contain graphics which use mathematical expressions that describe objects such as points, lines, curves and polygons, in order to generate an image. Vector images do not have a definite number of pixels as the objects are mathematical descriptions and can be geometrically transformed to any scale, rotation, translation etc.

Bandwidth is a measure of the width of a range of frequencies. It is the difference between the upper and lower frequencies occupied by a continuous set of frequencies. Bandwidth is a key determinant of the channel capacity. For example, assuming equivalent noise levels a 1kHz band can carry the same amount of data regardless of its operational frequency, therefore other methods are required to improve the channel capacity. One such method for reducing the bandwidth required is to use image compression techniques. Image compression addresses the problem of reducing the amount of data required to represent a digital image [104].

Superimposing the $wFoV$ with the $nFoV$ can be considered a form of image compression. The high spatial details found in the $nFoV$ imagery have been embedded into the $wFoV$ data. The information is then transmitted to its intended recipient i.e. remote terminal operator. In this section a simple superimposed video will be created using the standard MPEG-4 (Moving Picture Experts Group) format. This format has been chosen as it is the most commonly used format to store digital video and audio data and is an open international standard multimedia framework [94, 133].

The aim of this study is to quantitatively analyse the reduction in bandwidth required to transmit two separate $wFoV$ and $nFoV$ image data streams as opposed to a single stream of the superimposed image data. Two $nFoV$ image data sets were tested. Firstly, the $nFoV$ video data captured from a compact digital camera (Nikon Coolpix L27). Secondly an augmented version of the $nFoV$ data was tested. The augmented video data is acquired from the captured $wFoV$ video data. The augmented $nFoV$ is used to show that the high frequency spatial details are still present in the superimposed video, in comparison to the augmented superimposed video, and to show the difference in bandwidth savings between both. The videos are originally in the AVI format using Motion JPEG video encoding. Firstly, the augmented $nFoV$ stream was created using the best quality available and keeping the encoding the same, to remove any further compression. Then similarly, the superimposed video streams were created. The bandwidth savings were calculated using the total sum of the $wFoV$ and $nFoV$ at different stages (frames) in the video.

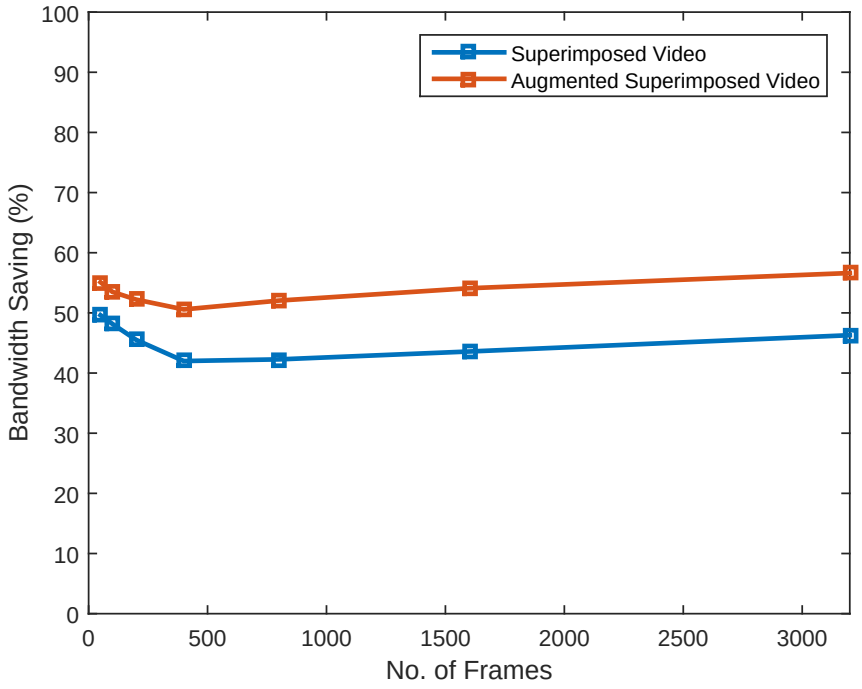


Figure A.1: Illustration of the bandwidth savings (%) with variation to the number of image frames (video)

The superimposed video generated from the captured *wFoV* and *nFoV* videos shows the potential bandwidth savings available for use with the novel imaging system. The augmented video shows higher bandwidth reductions but this is most likely due to the lower spatial details present in the augmented superimposed video, as compared to the captured *nFoV* video. The results show that utilising the geometric separation method to transport both the *wFoV* and *nFoV* streams, in the superimposed manner, can provide at minimum, a 40% reduction in bandwidth.

A.2 MPEG Compression

The previous section on bandwidth savings, led to the research of how well the MPEG-4 standard is able to compress the video streams to further reduce the stream size, whilst minimising the loss in visual detail. The test in the previous section was re-conducted with different levels of compression. The file sizes of the data streams were once again obtained at different stages of the video production but additionally the structural similarity (SSIM) index [276] was used to assess the image quality. This was due to its accuracy, robustness, availability and also due to its wide use as an industry standard image quality metric (IQM). Figures A.2 and A.3 compares the file sizes and the image quality at different levels of compression.

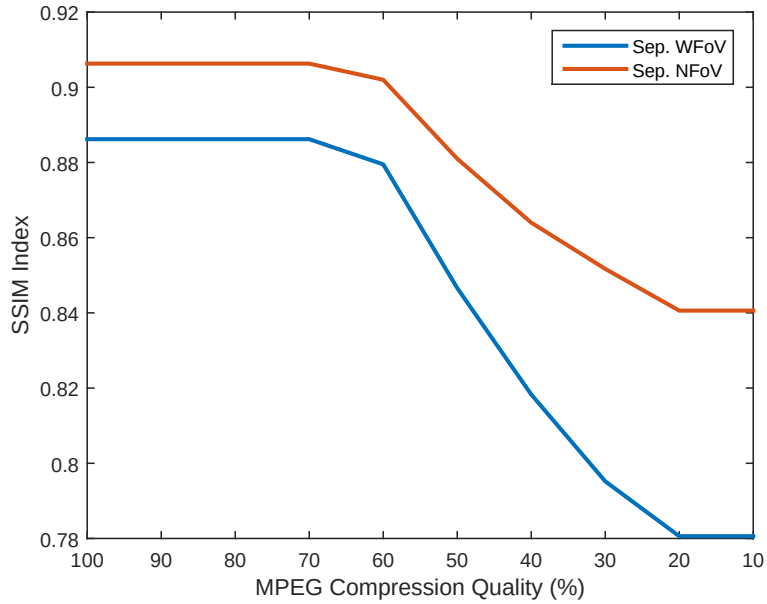


Figure A.2: Illustration of variation of image quality using SSIM as the MPEG compression quality varies.

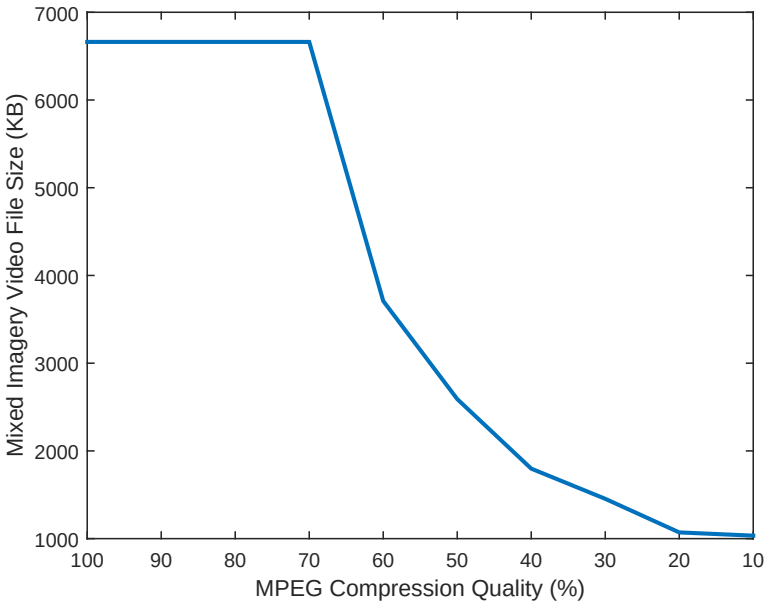
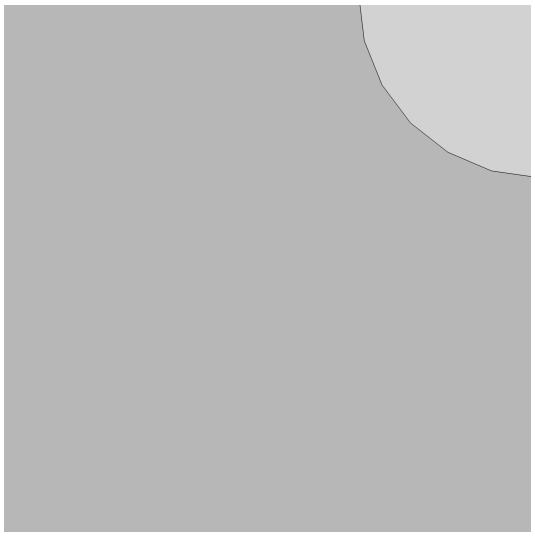


Figure A.3: Illustration of variation of video file size (KB) of superimposed imagery as the MPEG compression quality varies.

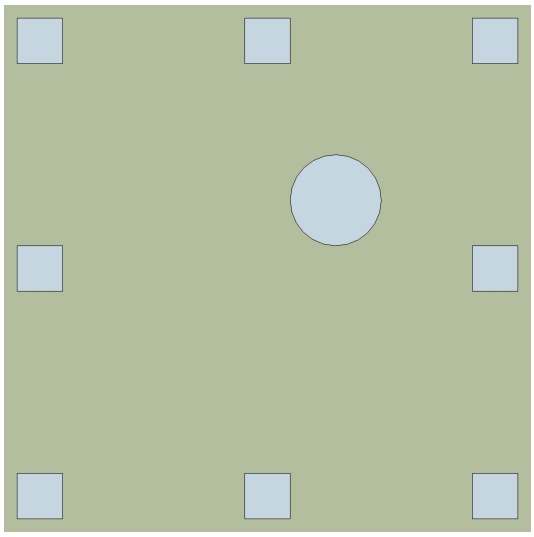
The results in Figure A.2 indicate that below 60% compression quality the image quality significantly reduces. Figure A.3 shows a similar trend. The key finding from the figures show that at 60% quality, there is a noticeable saving in the filesize, which can translate to a reduction in the stream size but, as shown in Figure A.2, there is only a small reduction in quality (a reduction of less than 1%).

Appendix B

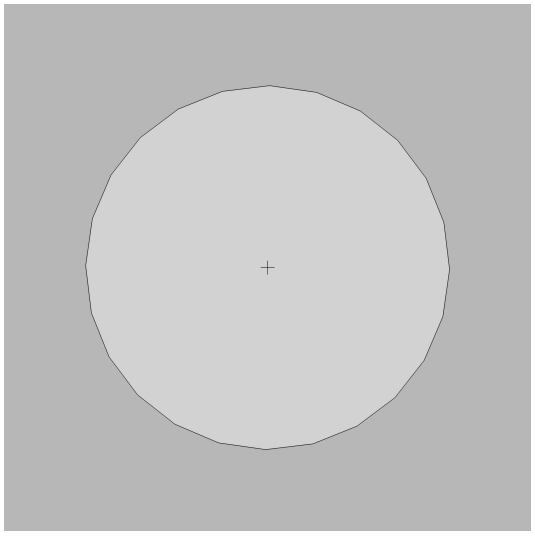
Image Database



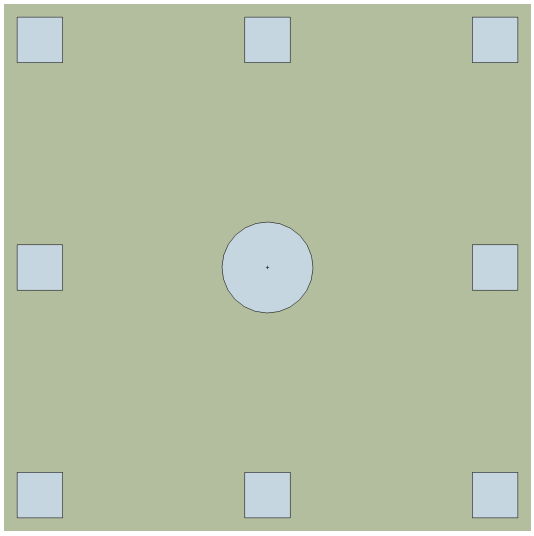
(1) Self-created vector image ‘BGO[nfov].jpg’



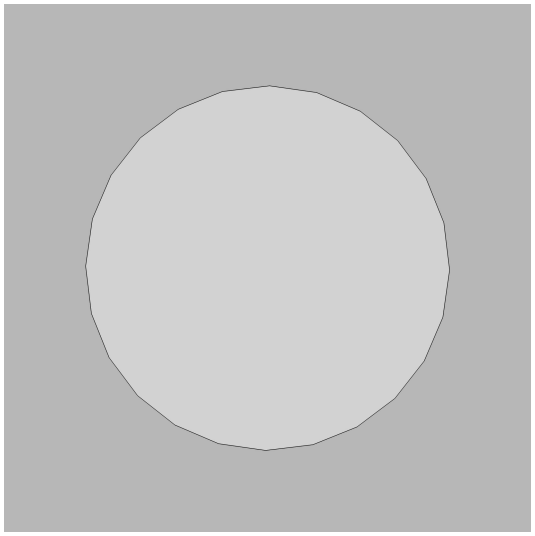
(2) Self-created vector image ‘BGO[wfov].jpg’



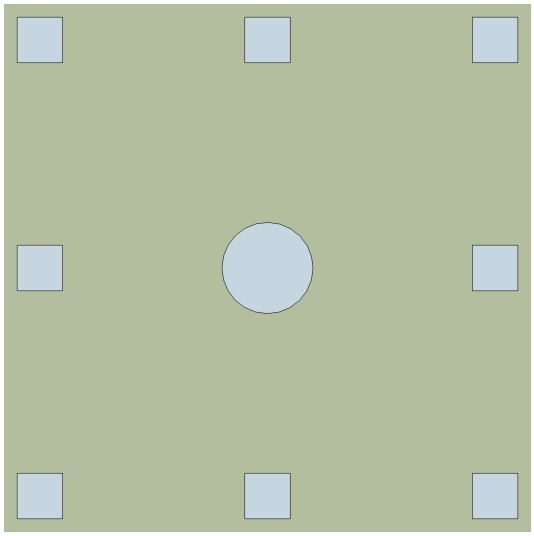
(3) Vector image ‘BGO_000[nfov].jpg’



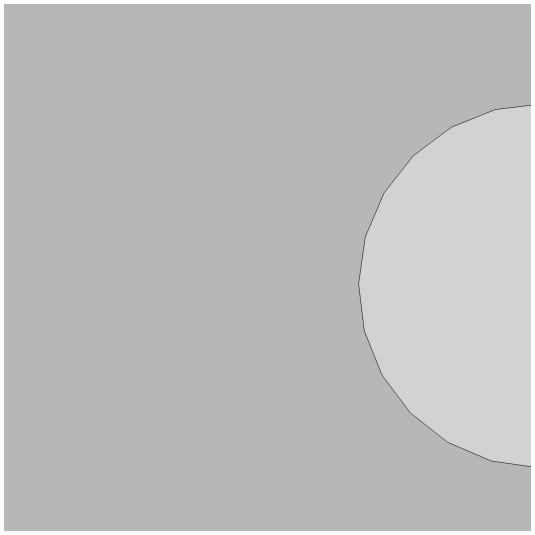
(4) Vector image ‘BGO_000[wfov].jpg’



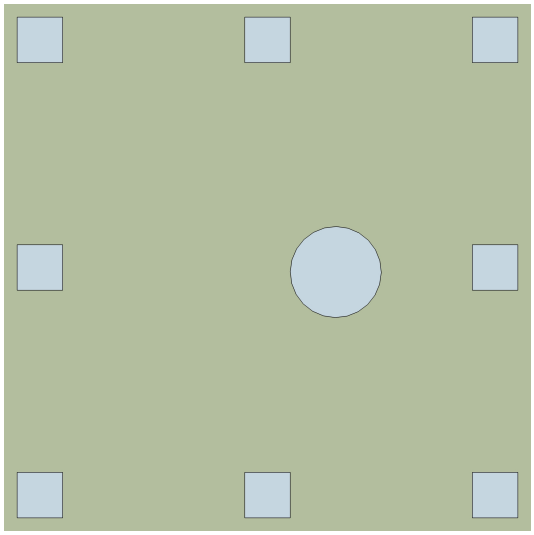
(5) Vector image ‘BGO_00[nfov].jpg’



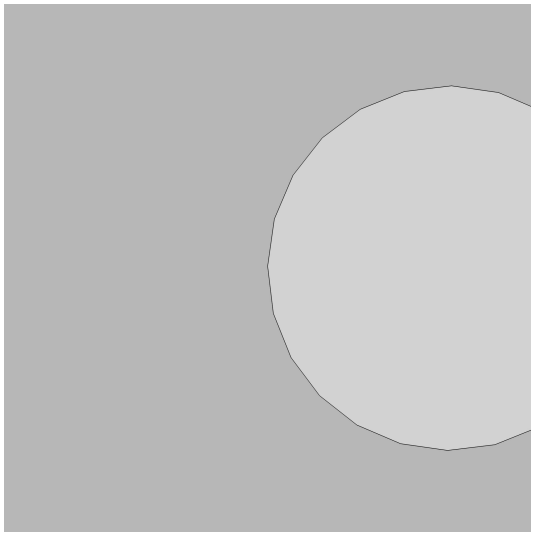
(6) Vector image ‘BGO_00[wfov].jpg’



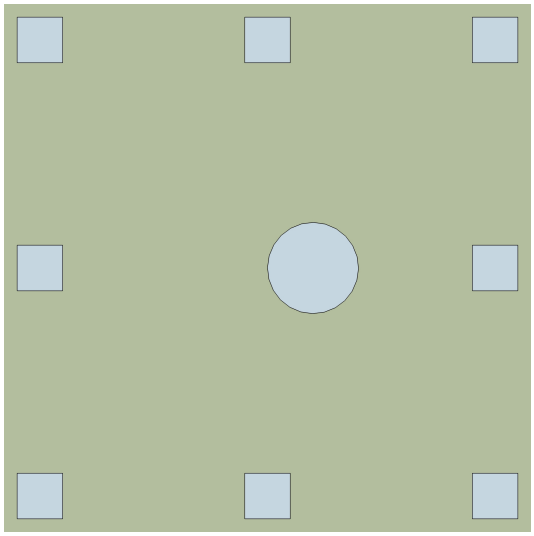
(7) Vector image ‘BGO_2[nfov].jpg’



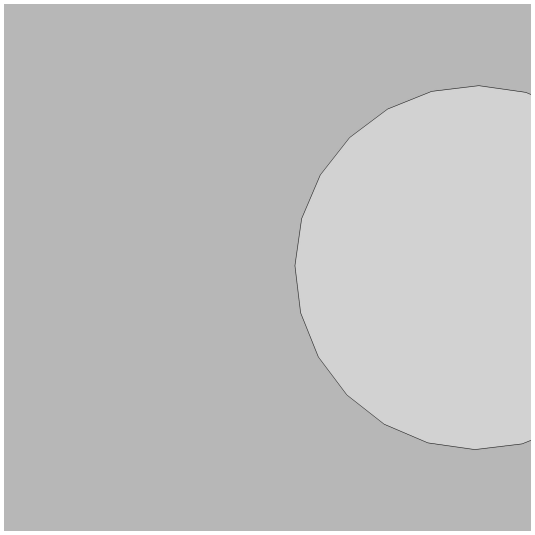
(8) Vector image ‘BGO_2[wfov].jpg’



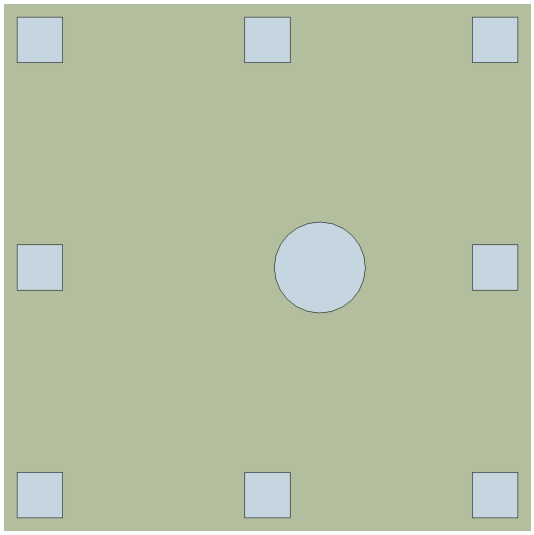
(9) Vector image ‘BGO_3[nfov].jpg’



(10) Vector image ‘BGO_3[wfov].jpg’



(11) Vector image ‘BGO_5[nfov].jpg’



(12) Vector image ‘BGO_5[wfov].jpg’

(13) Online, taken using Olympus DSLR 'Brisbane.tiff'

(14) Online, vector image, uncompressed, 'Cartoon Landscapes Vector Background.tif'

(15) Online, vector image, uncompressed, 'Chicago.tiff'

(16) Self-captured using Casio DSLR, uncompressed, 'Croatia.tiff'

(17) Self-captured using Nikon Coolpix L27 point-and-shoot camera 'DSCN0008.jpg'

(18) Self-captured using Nikon Coolpix L27 camera 'DSCN0022.jpg'

(19) Self-captured using Nikon Coolpix L27 camera 'DSCN0026.jpg'

(20) Self-captured using Nikon Coolpix L27 camera 'DSCN0039.jpg'

(21) Self-captured using Nikon Coolpix L27 camera 'DSCN0126.jpg'

(22) Self-captured using Nikon Coolpix L27 camera 'DSCN0144.jpg'

(23) Self-captured using Nikon Coolpix L27 camera 'DSCN0177.jpg'

(24) Self-captured using Nikon Coolpix L27 camera 'DSCN0211.jpg'

(25) Self-captured using Nikon Coolpix L27 camera 'DSCN0237.jpg'

(26) Self-captured using Casio DSLR, uncompressed, 'Edinburgh1.tiff'

(27) Online, vector image, compressed, 'Landscapes-bf1-[Converted-noninterlaced].png'

(28) Online, vector image, compressed, 'Night-cityscape.png'

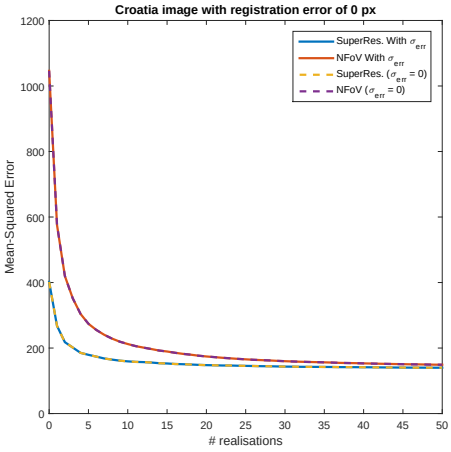
(29) Self-captured using Casio DSLR, compressed, ‘Queenstown1.jpg’

(30) Self-captured using Casio DSLR, compressed, ‘Queenstown2.jpg’

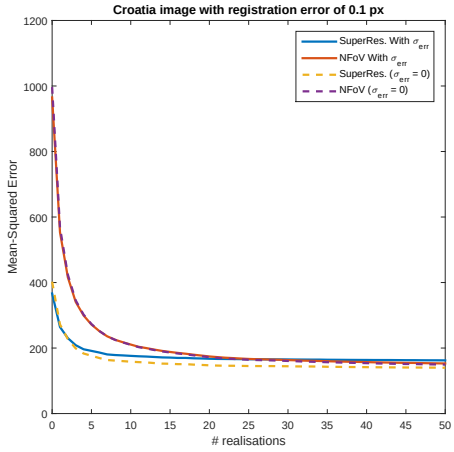
(31) Online, vector image, compressed, ‘Starlight-glow-Background-[Converted].png’

Appendix C

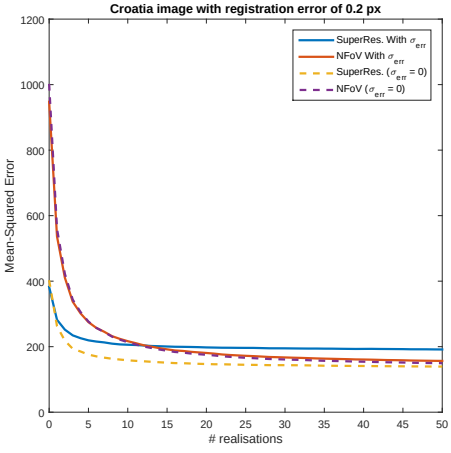
Temporal Averaging Separation Figures


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

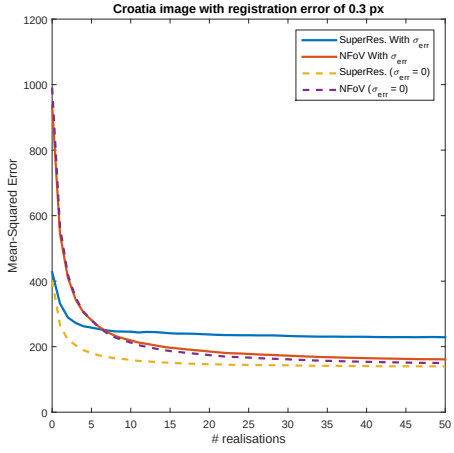
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

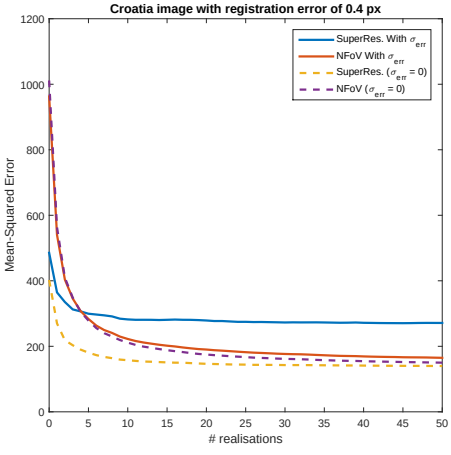
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

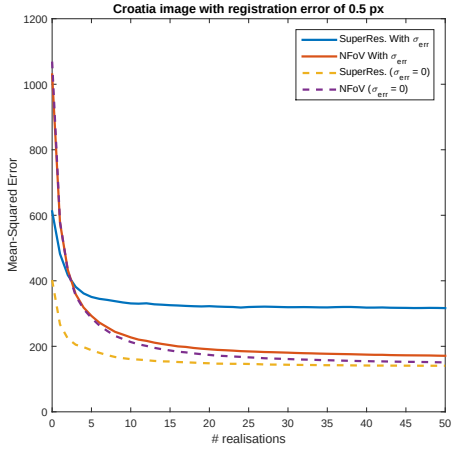
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

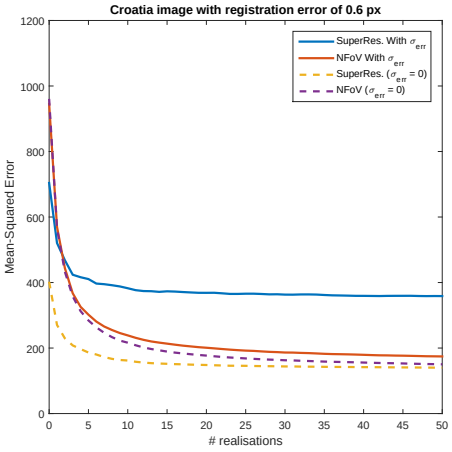
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

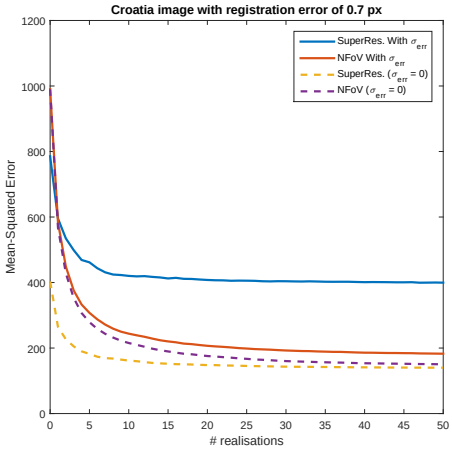
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

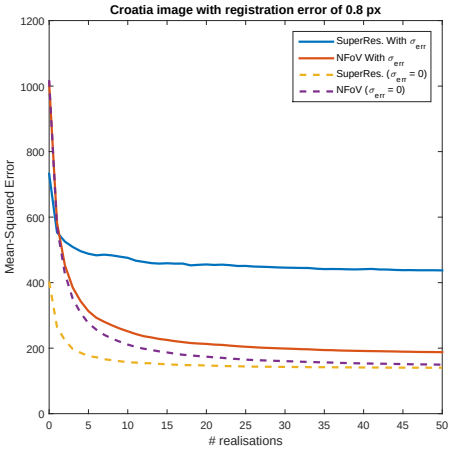
(f)



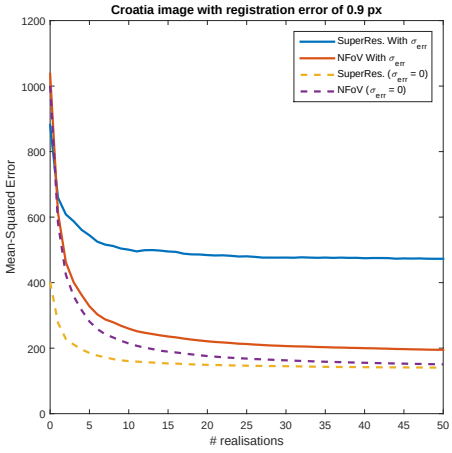
(g)



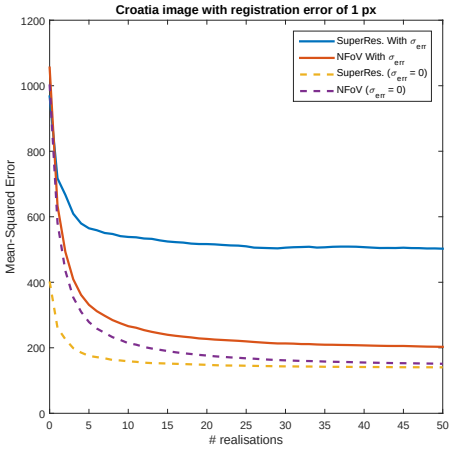
(h)



(i)

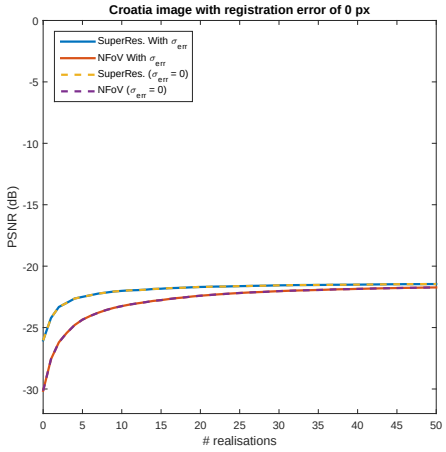


(j)

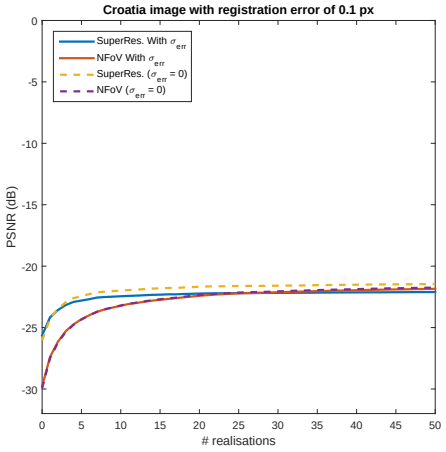


(k)

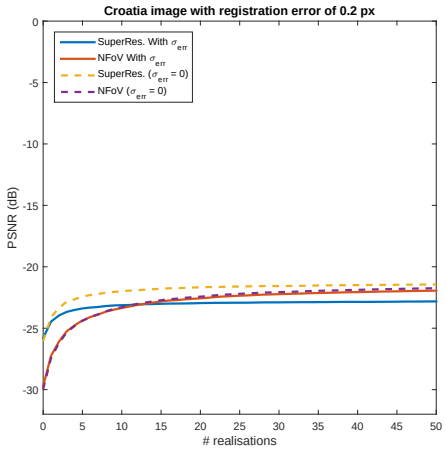
Figure C.1: Mean-squared error (MSE) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.1(a) - C.1(k))


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

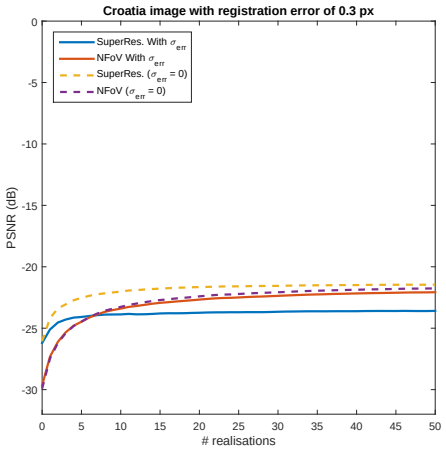
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

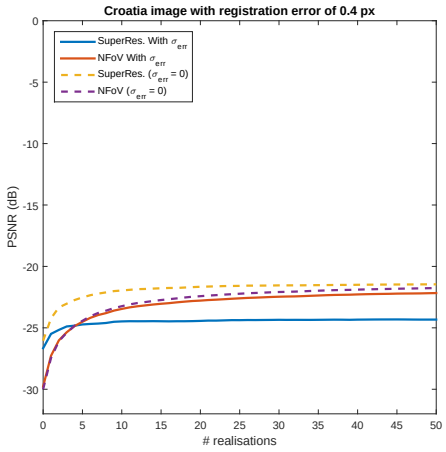
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

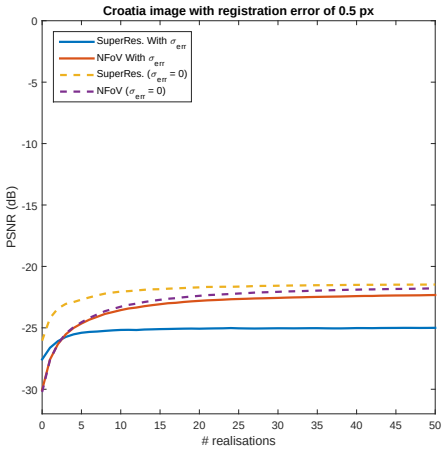
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

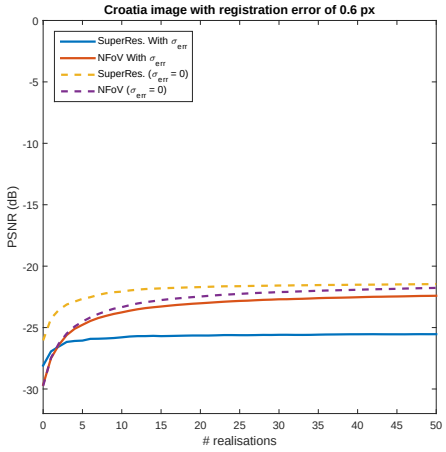
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

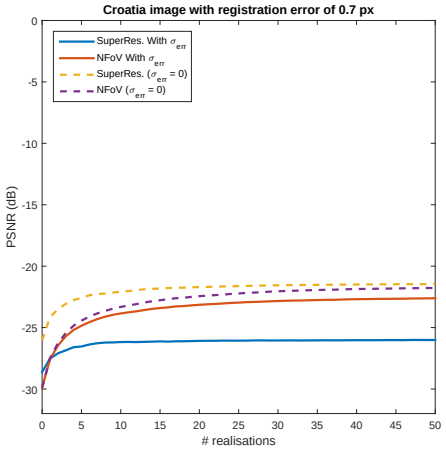
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

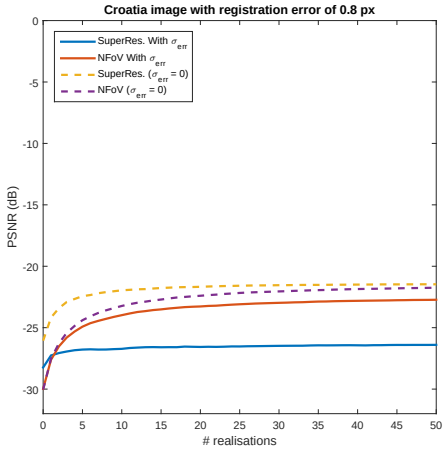
(f)



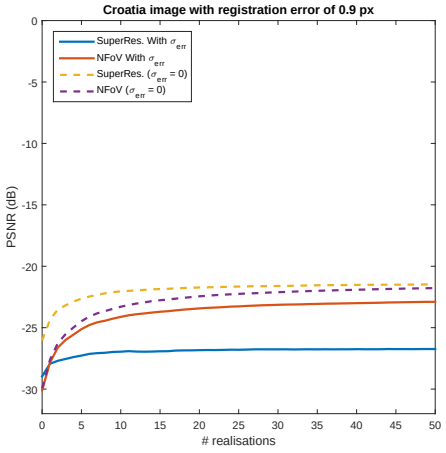
(g)



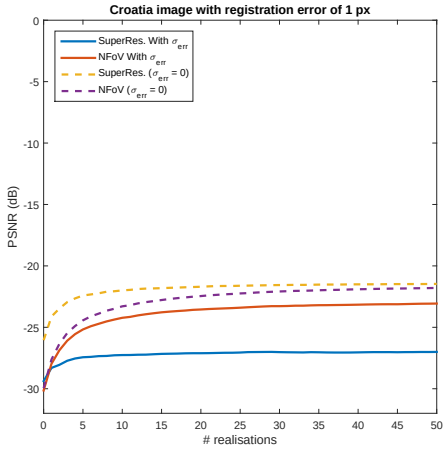
(h)



(i)

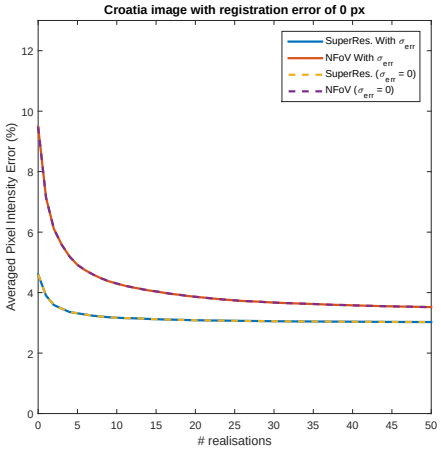


(j)

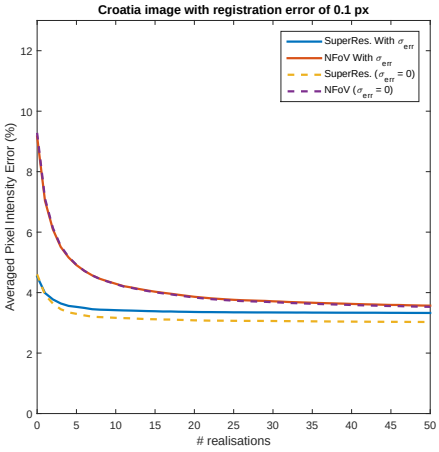


(k)

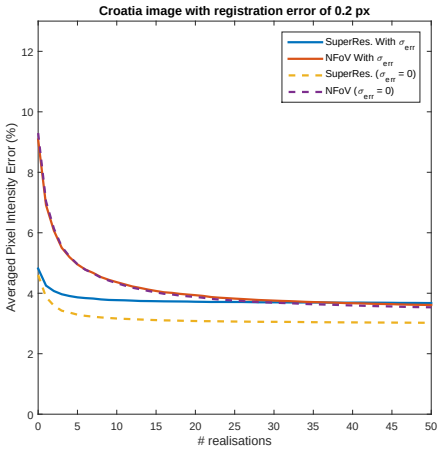
Figure C.2: Peak signal-to-noise ratio (PSNR) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.2(a) - C.2(k))


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

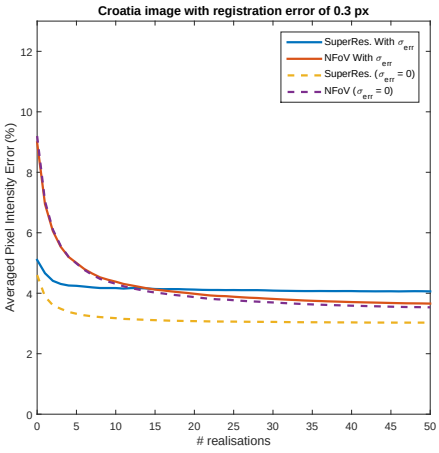
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

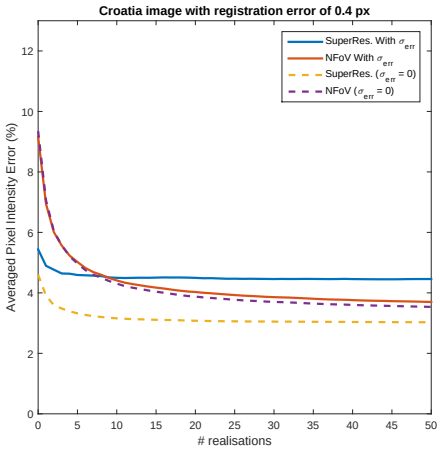
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

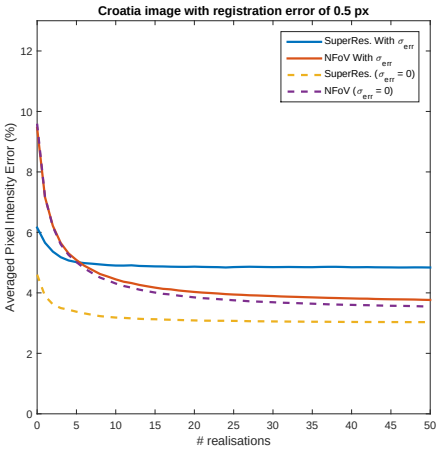
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

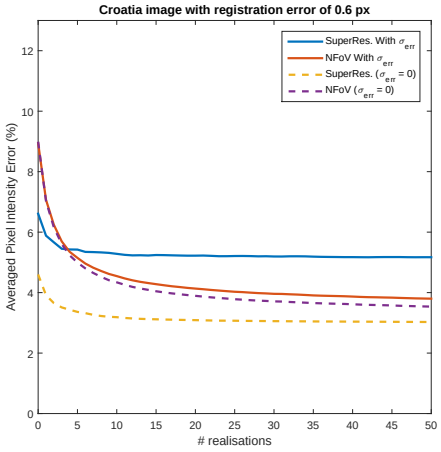
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

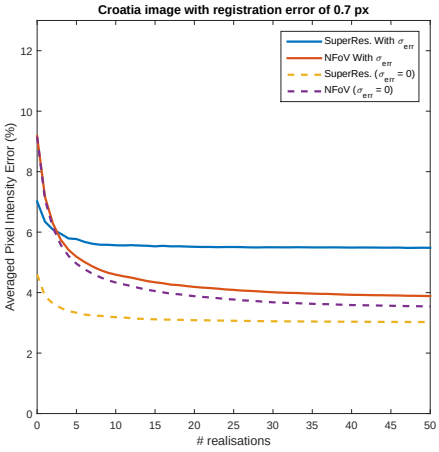
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

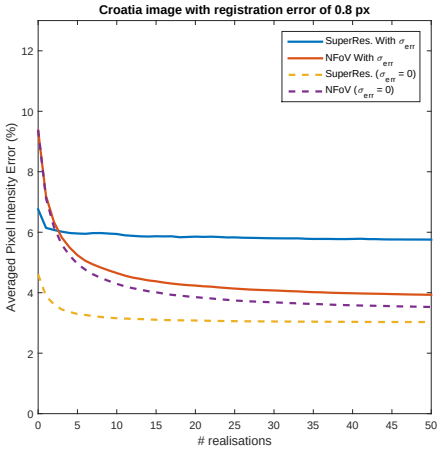
(f)



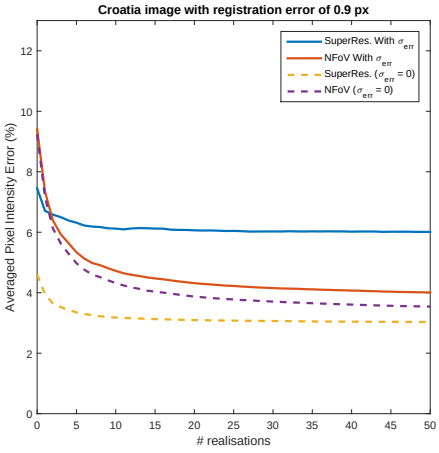
(g)



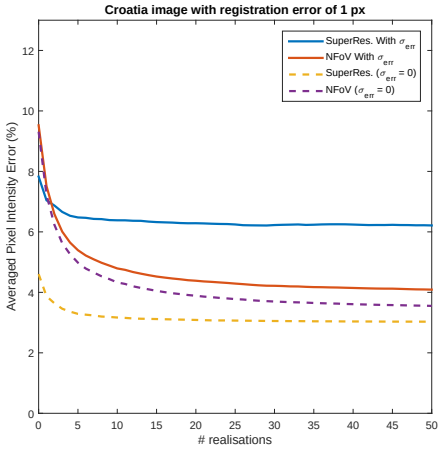
(h)



(i)

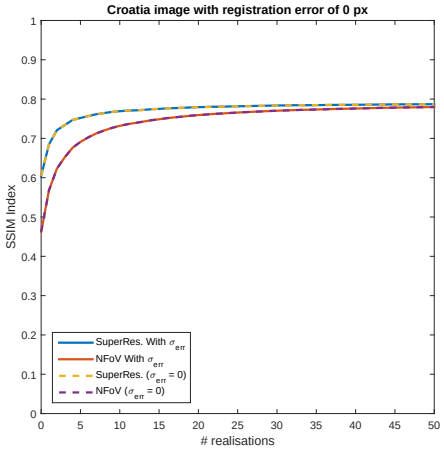


(j)

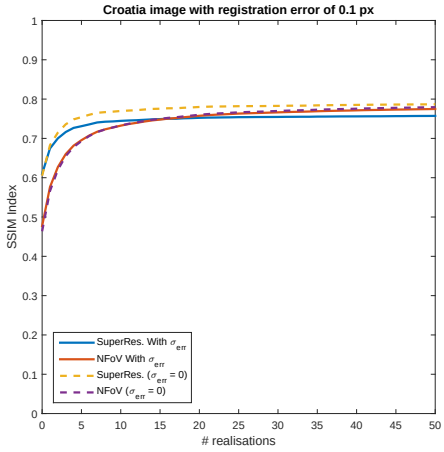


(k)

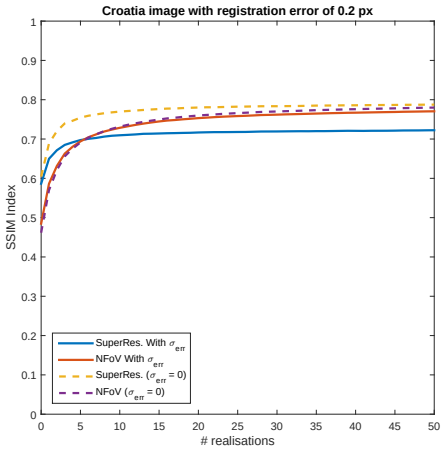
Figure C.3: Averaged pixel intensity error against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.3(a) - C.3(k))


Croatia image with registration error of 0 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

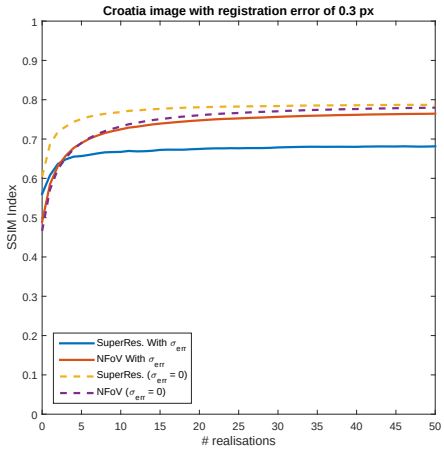
(a)


Croatia image with registration error of 0.1 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

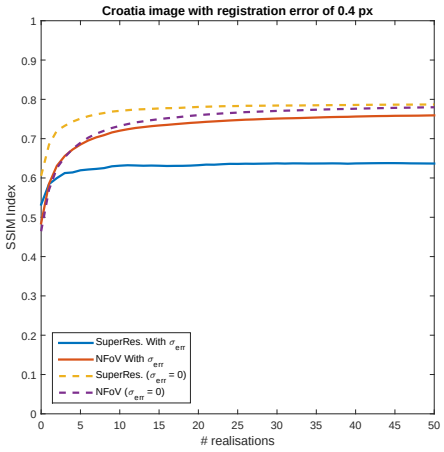
(b)


Croatia image with registration error of 0.2 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

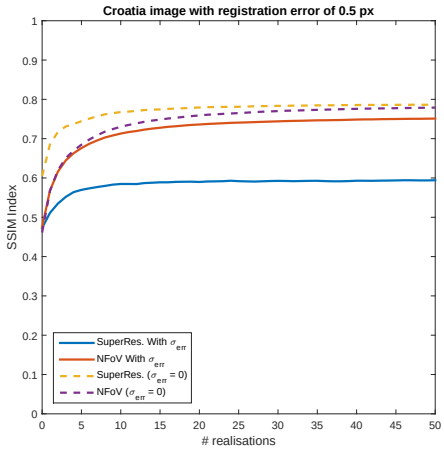
(c)


Croatia image with registration error of 0.3 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

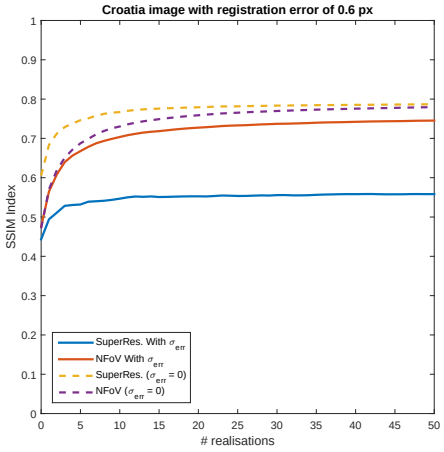
(d)


Croatia image with registration error of 0.4 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

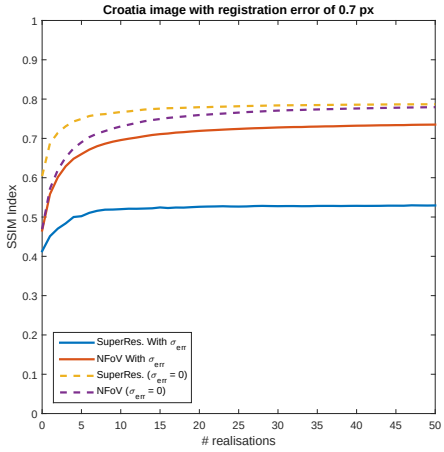
(e)


Croatia image with registration error of 0.5 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

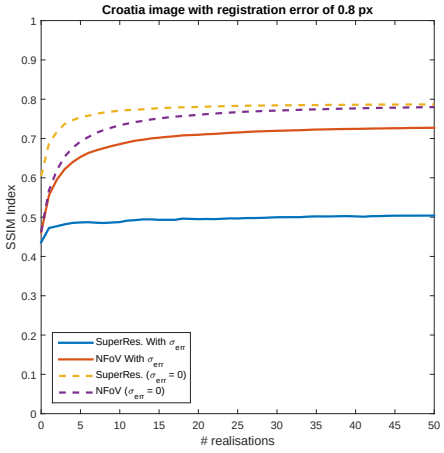
(f)



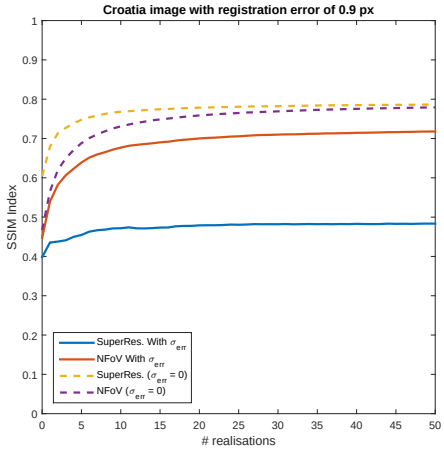
(g)



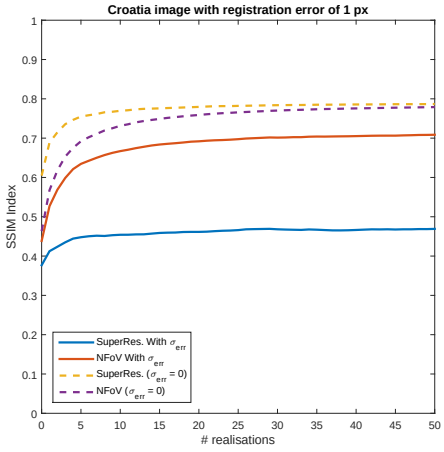
(h)



(i)

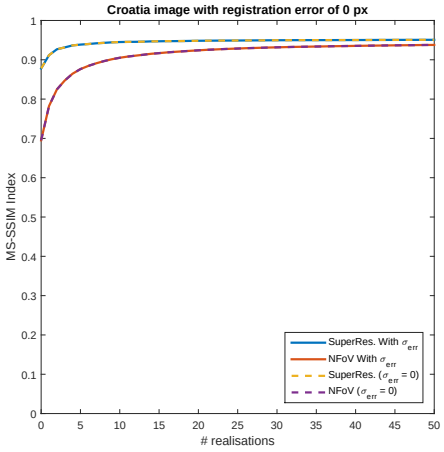


(j)

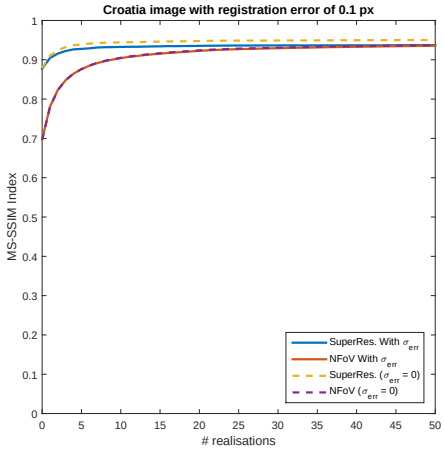


(k)

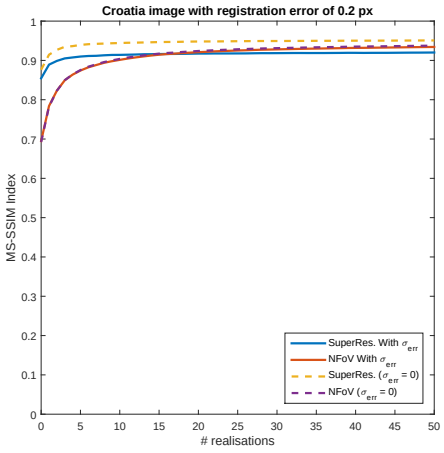
Figure C.4: Structural similarity (SSIM) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.4(a) - C.4(k))


Croatia image with registration error of 0 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

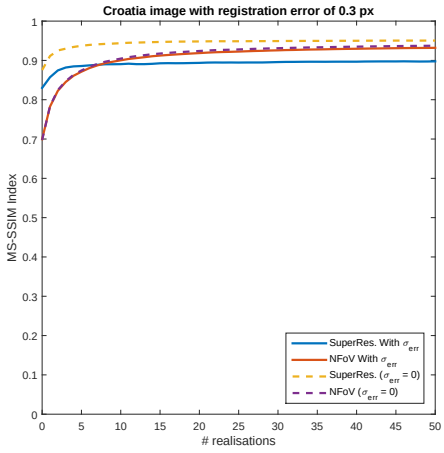
(a)


Croatia image with registration error of 0.1 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

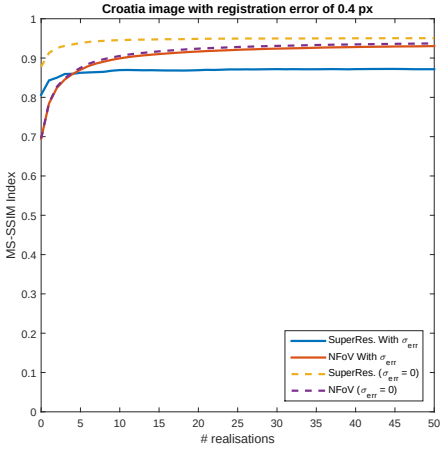
(b)


Croatia image with registration error of 0.2 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

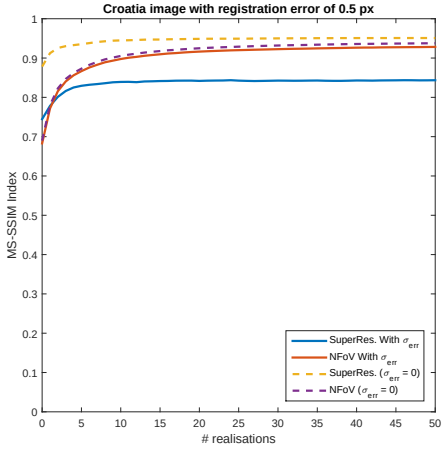
(c)


Croatia image with registration error of 0.3 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

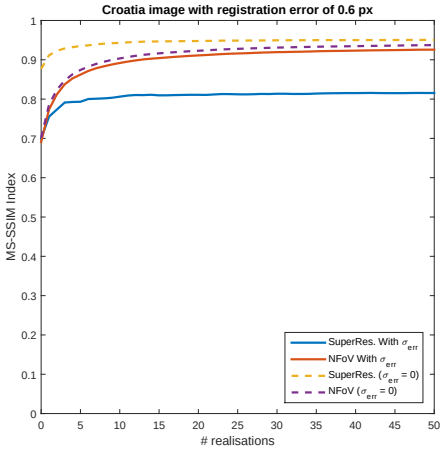
(d)


Croatia image with registration error of 0.4 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

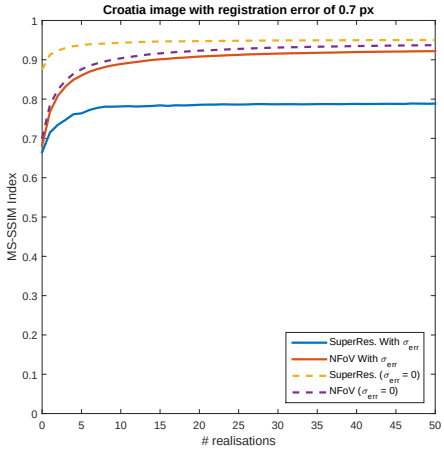
(e)


Croatia image with registration error of 0.5 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

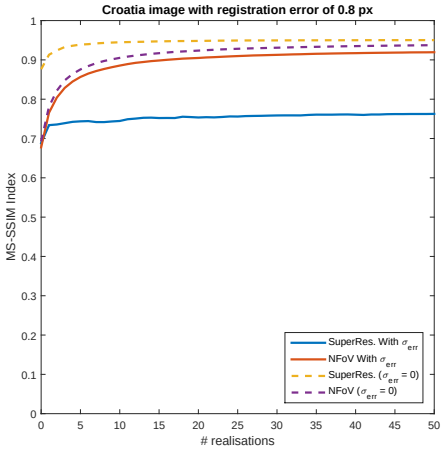
(f)



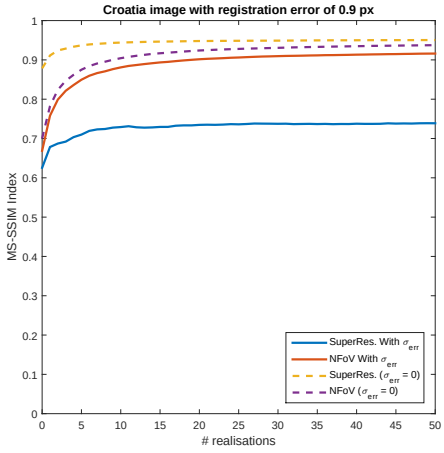
(g)



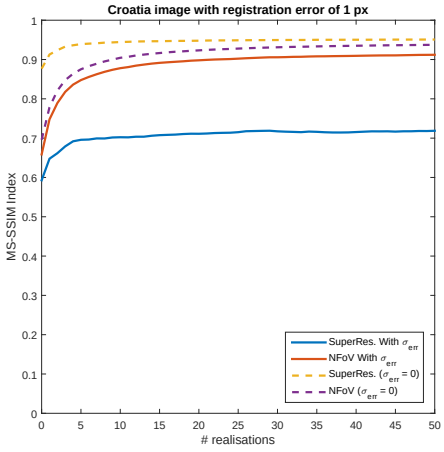
(h)



(i)

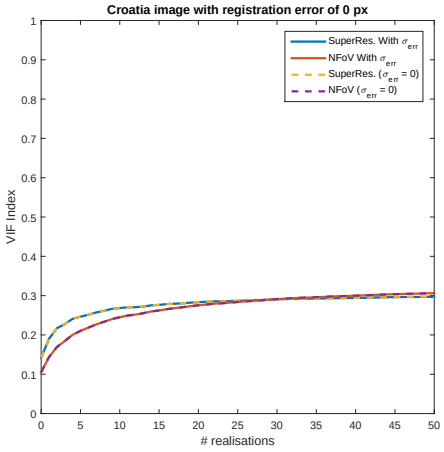


(j)

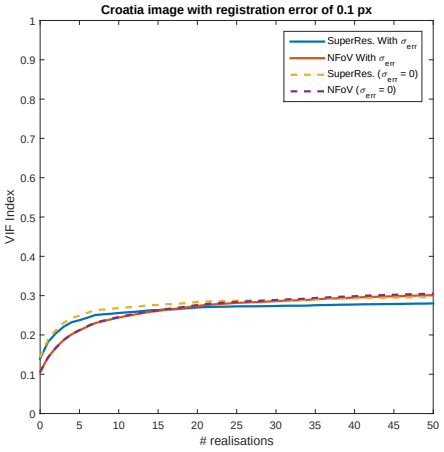


(k)

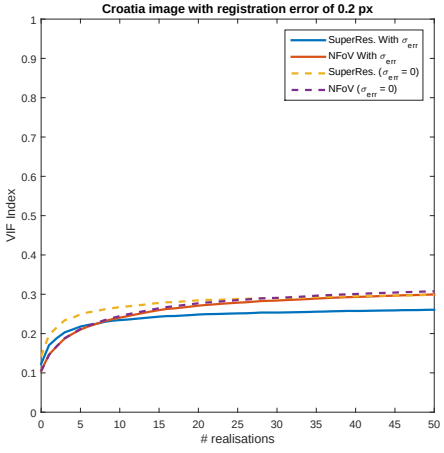
Figure C.5: Multi-scale structural similarity (MS-SSIM) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.5(a) - C.5(k))


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

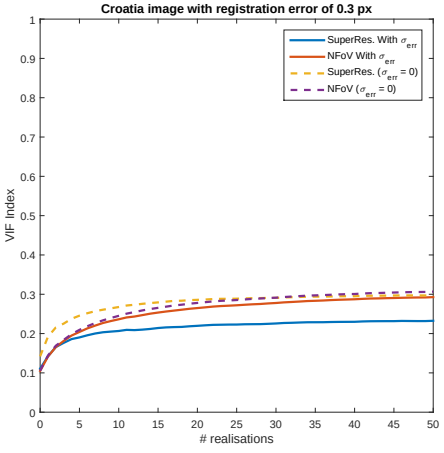
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

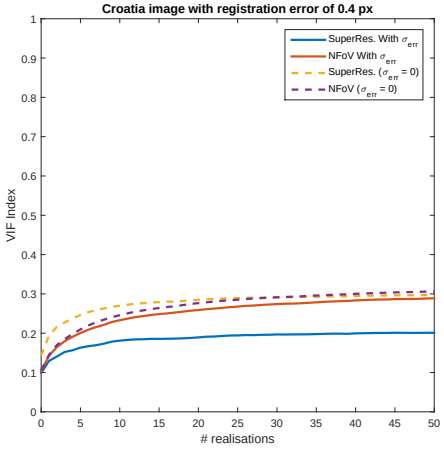
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

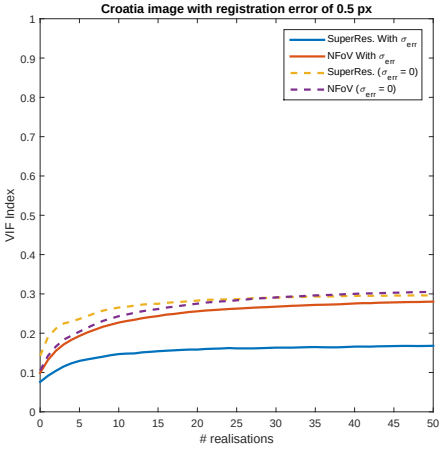
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

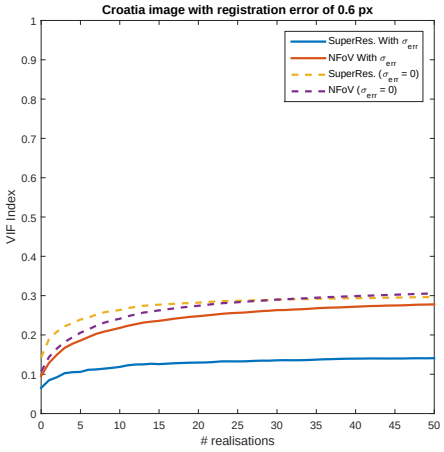
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

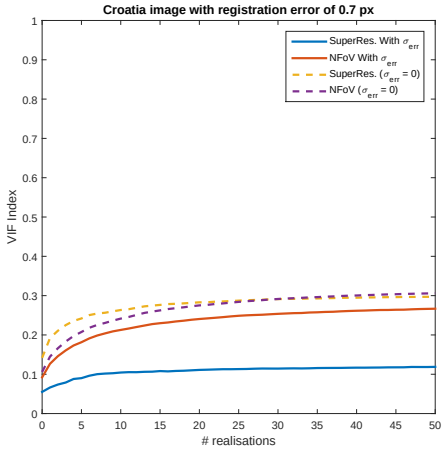
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

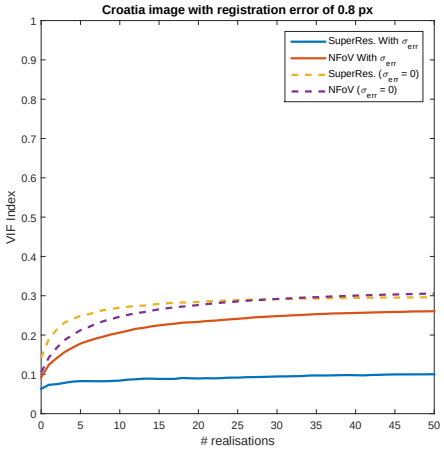
(f)



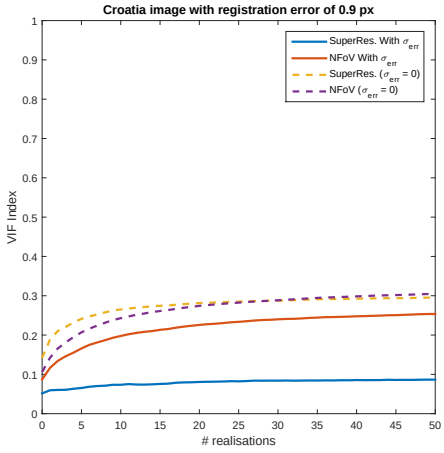
(g)



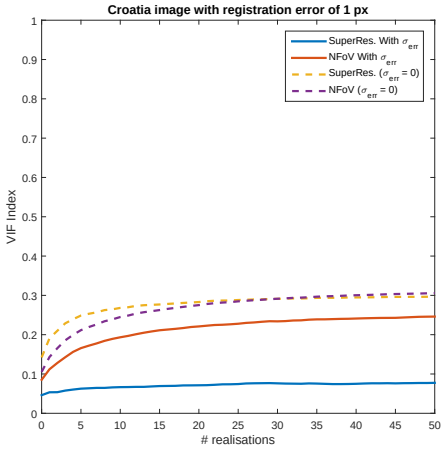
(h)



(i)

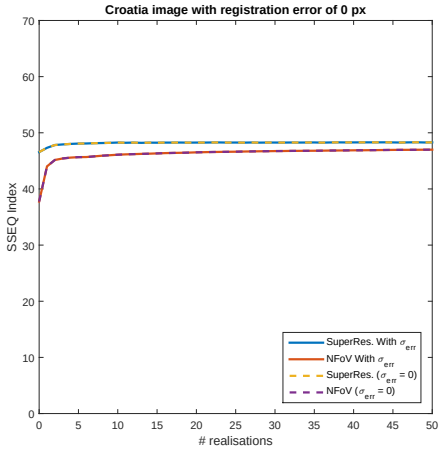


(j)

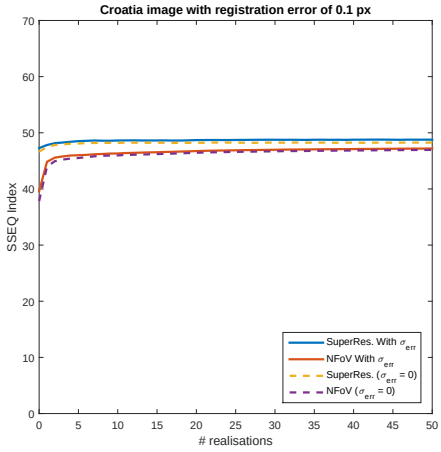


(k)

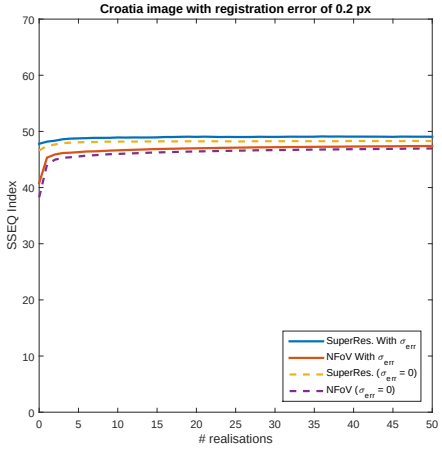
Figure C.6: Visual information fidelity (VIF) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.6(a) - C.6(k))


Croatia image with registration error of 0 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

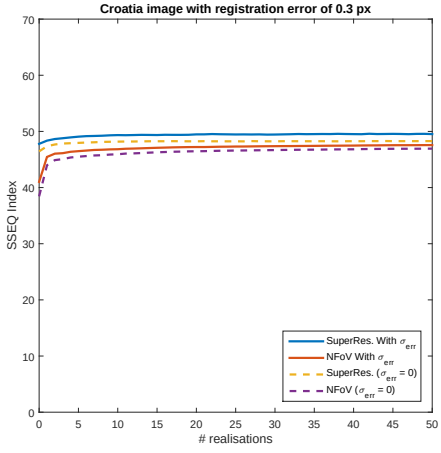
(a)


Croatia image with registration error of 0.1 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

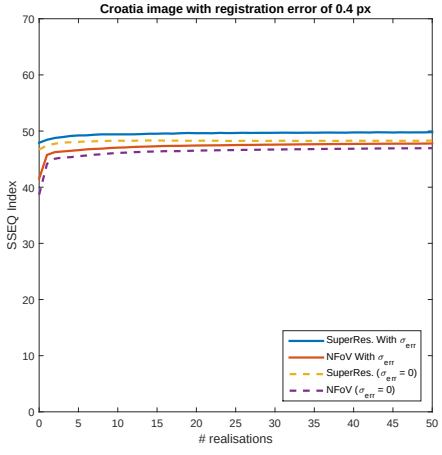
(b)


Croatia image with registration error of 0.2 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

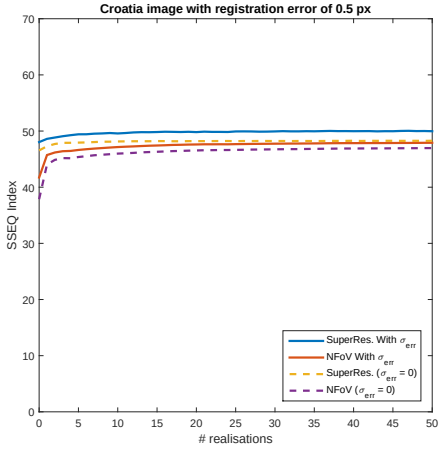
(c)


Croatia image with registration error of 0.3 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

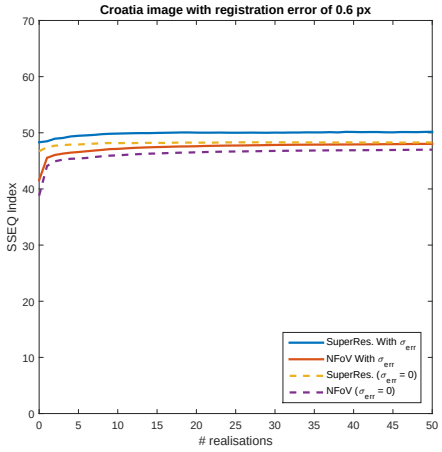
(d)


Croatia image with registration error of 0.4 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

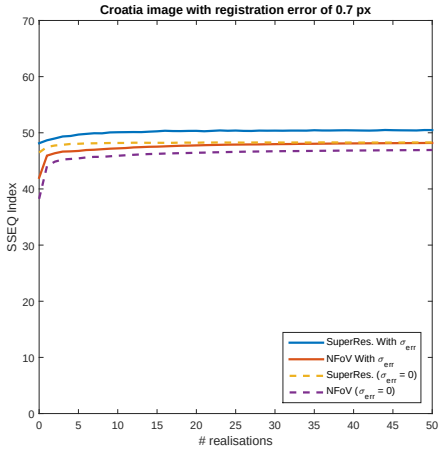
(e)


Croatia image with registration error of 0.5 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

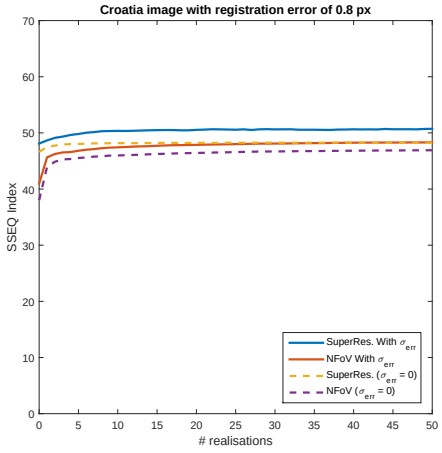
(f)



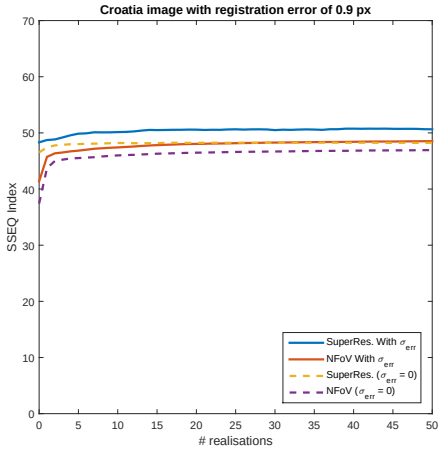
(g)



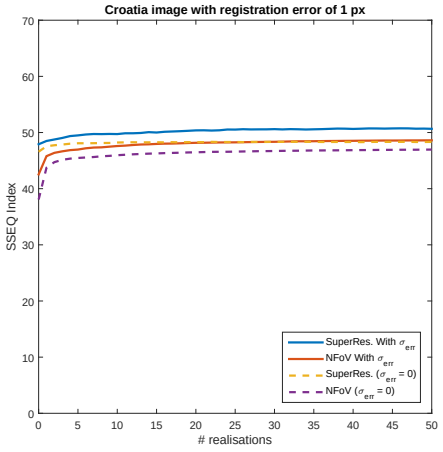
(h)



(i)

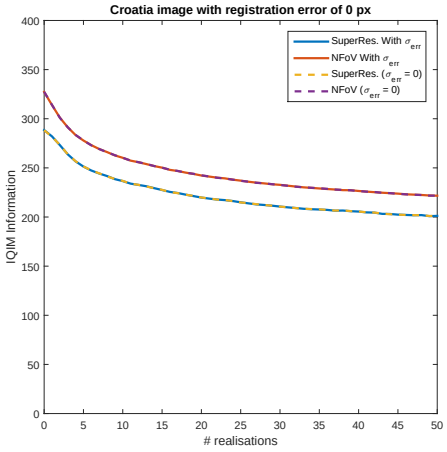


(j)

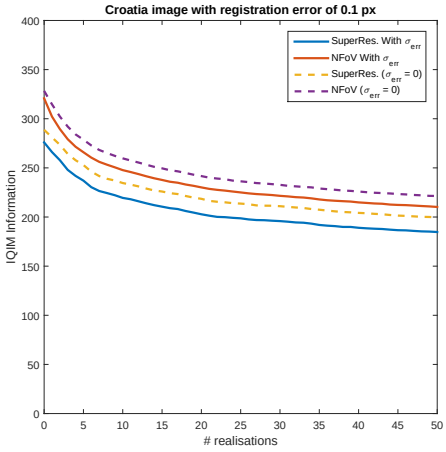


(k)

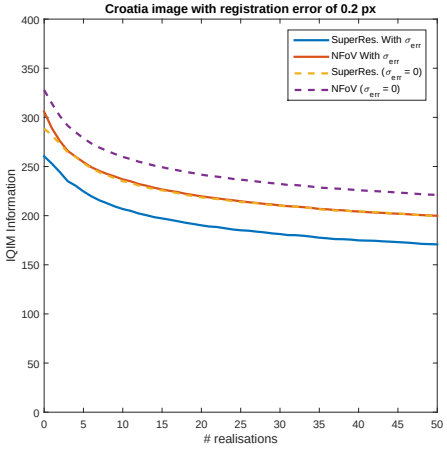
Figure C.7: Spatial-spectral entropy-based quality (SSEQ) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.7(a) - C.7(k))


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Information
realisations

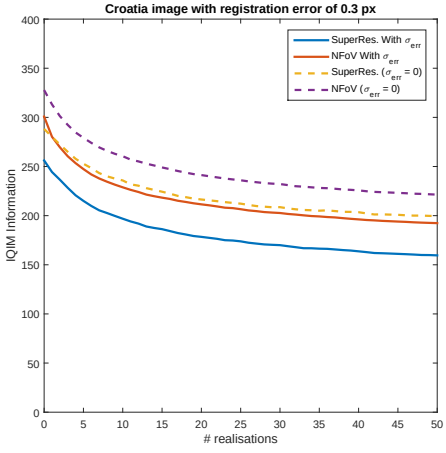
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Information
realisations

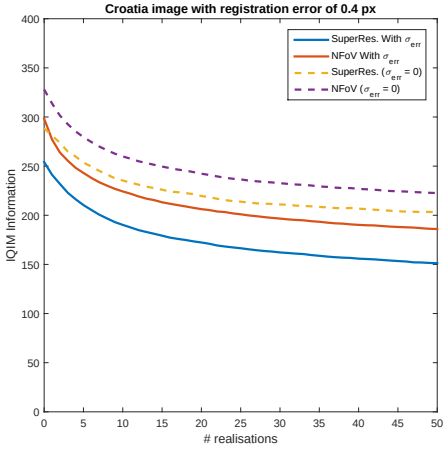
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Information
realisations

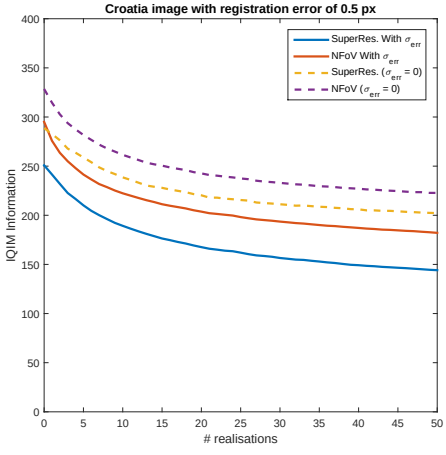
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Information
realisations

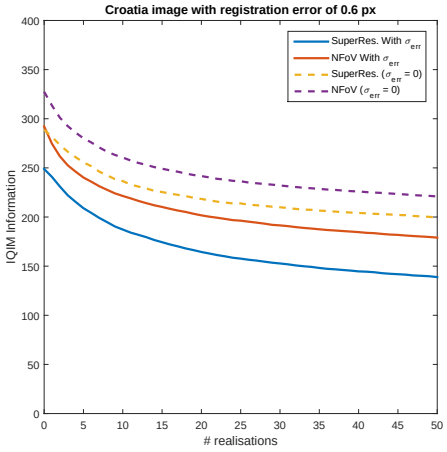
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Information
realisations

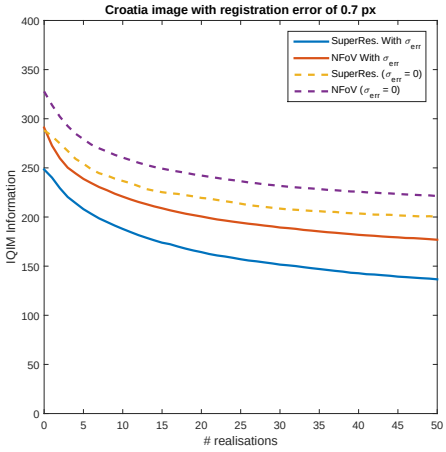
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Information
realisations

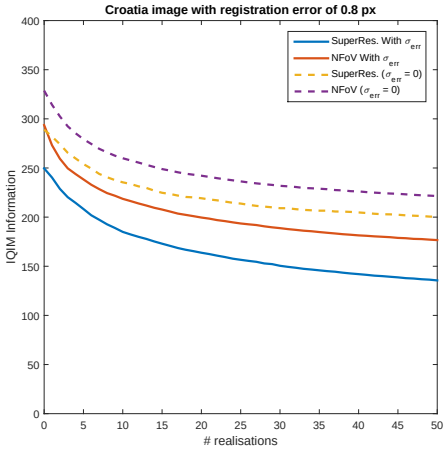
(f)



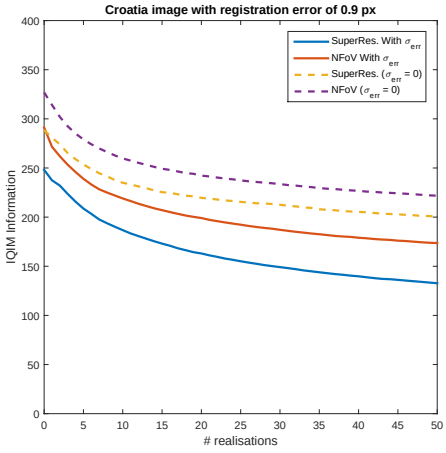
(g)



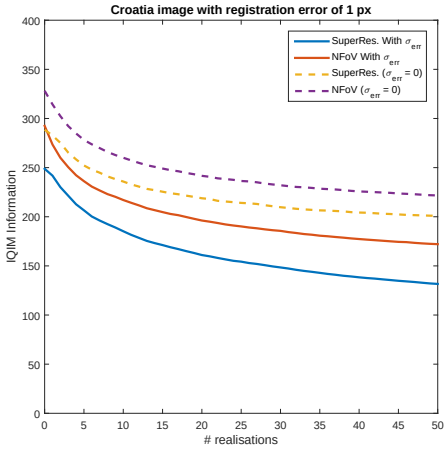
(h)



(i)

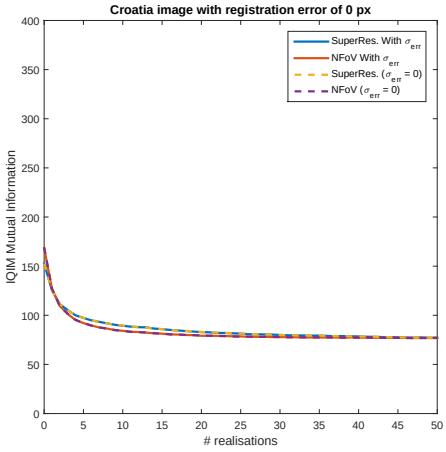


(j)

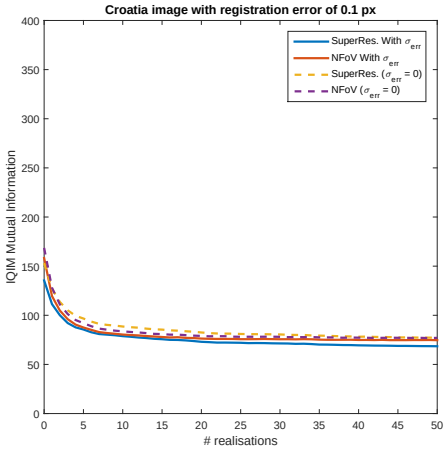


(k)

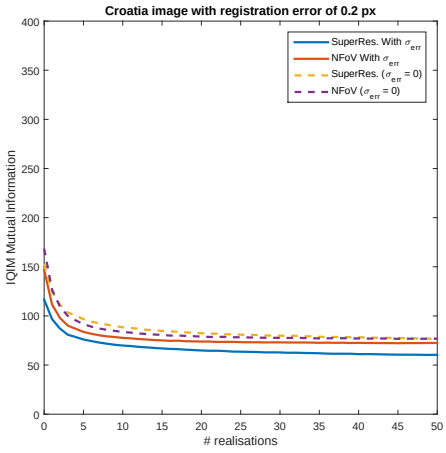
Figure C.8: Information quality information metric (IQIM) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.8(a) - C.8(k))


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

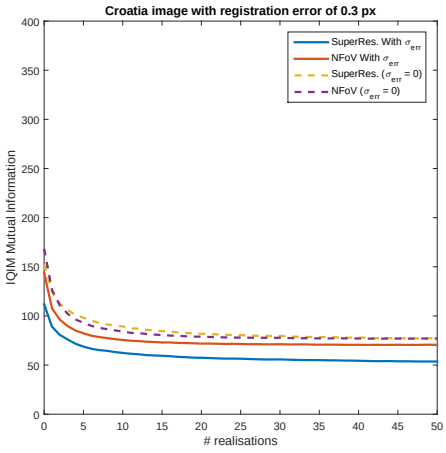
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

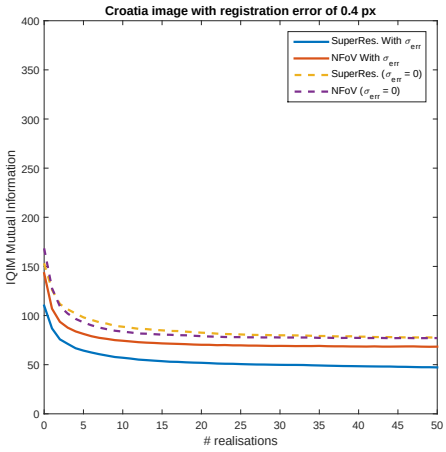
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

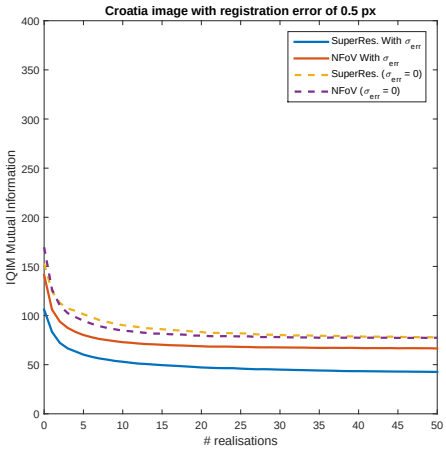
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

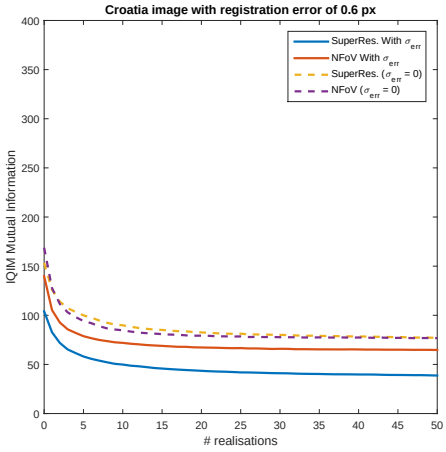
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

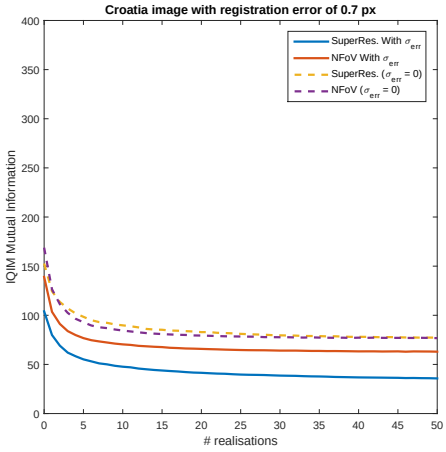
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

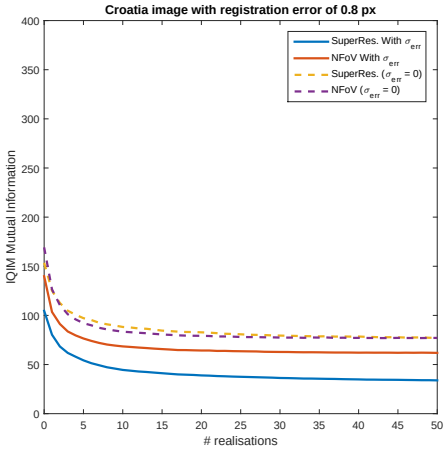
(f)



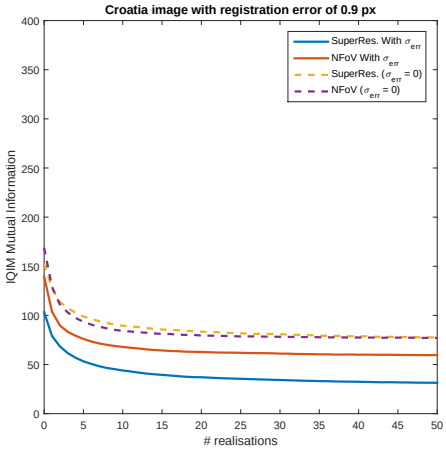
(g)



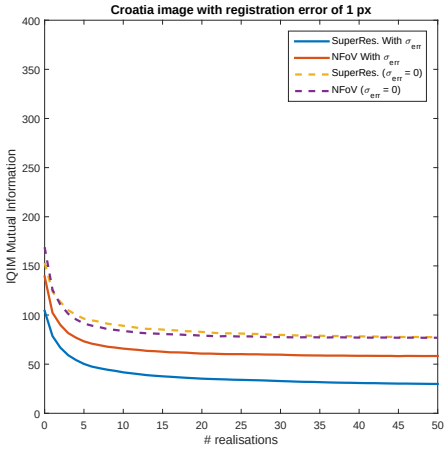
(h)



(i)

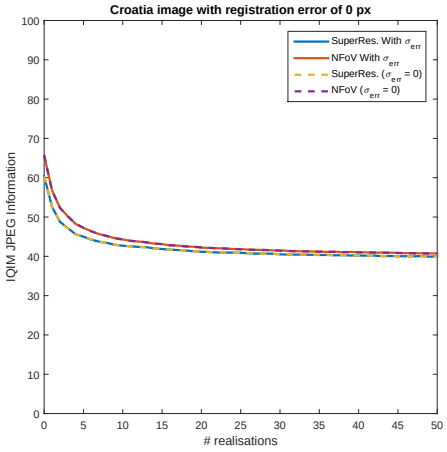


(j)

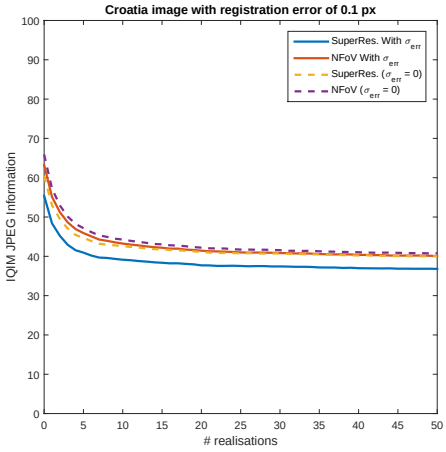


(k)

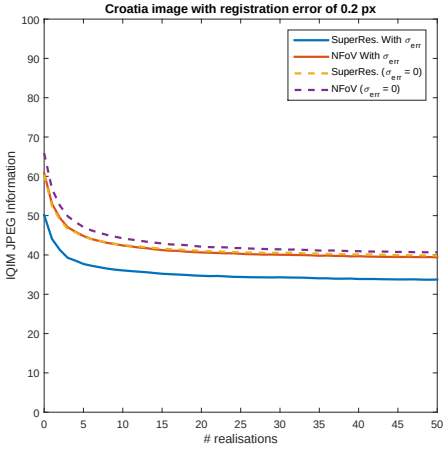
Figure C.9: Information quality information metric (mutual) (IQIMM) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.9(a) - C.9(k))


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

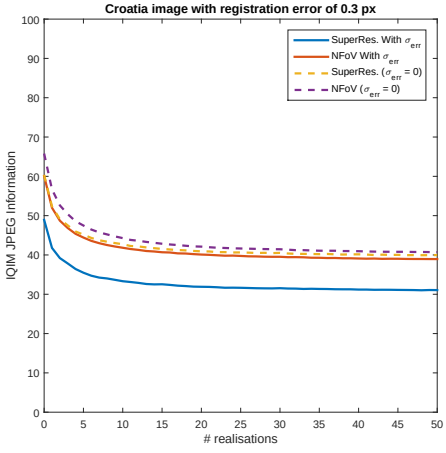
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

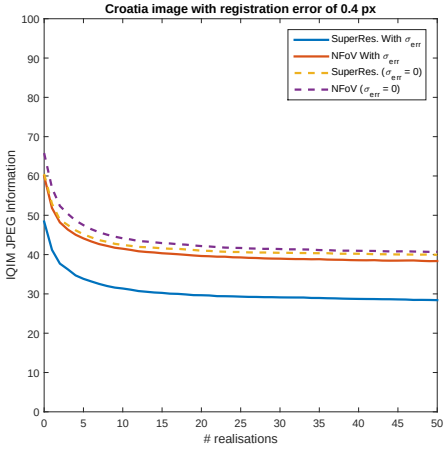
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

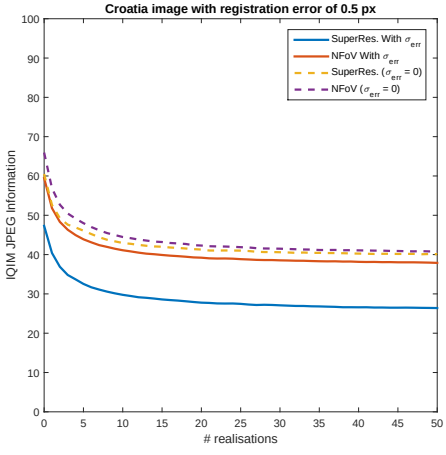
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

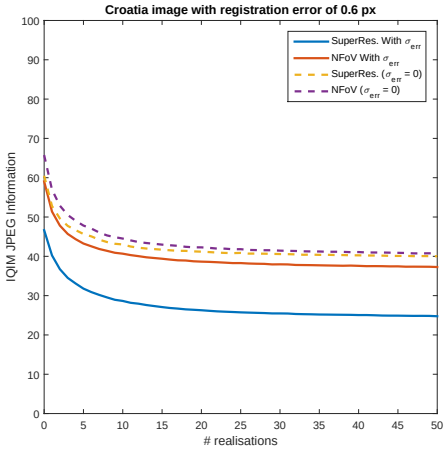
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

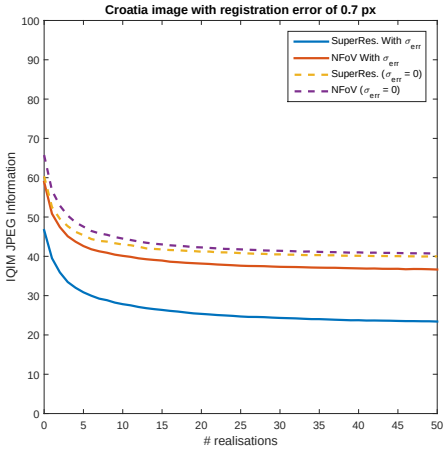
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

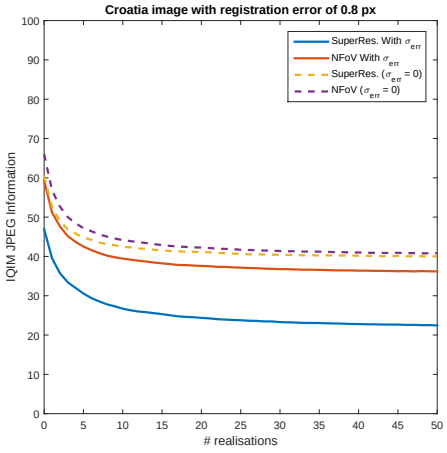
(f)



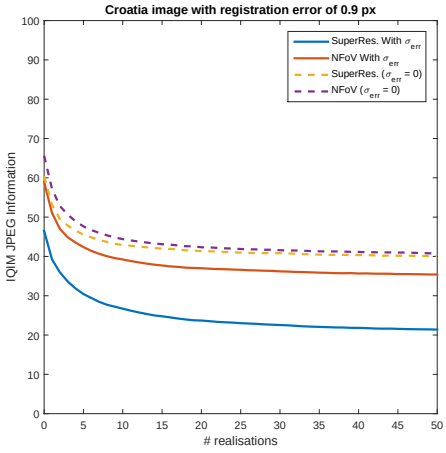
(g)



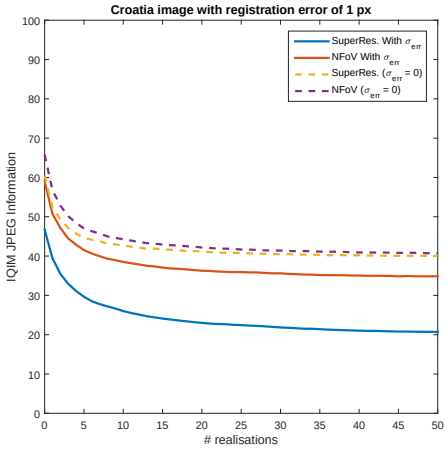
(h)



(i)

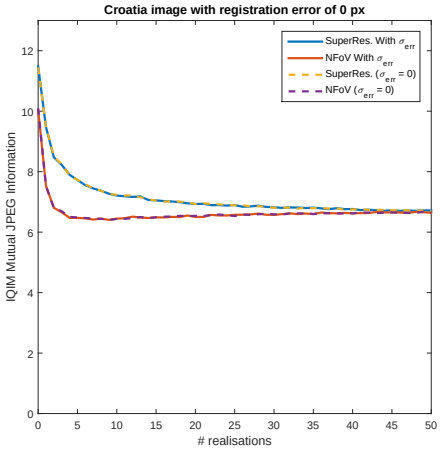


(j)

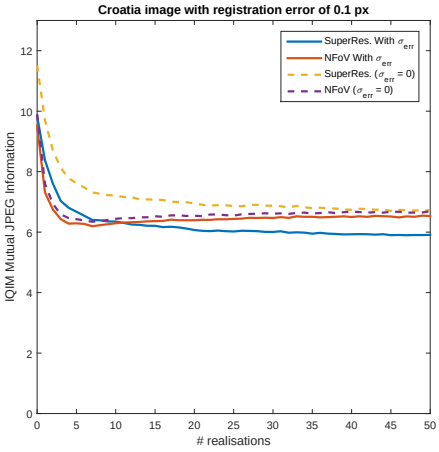


(k)

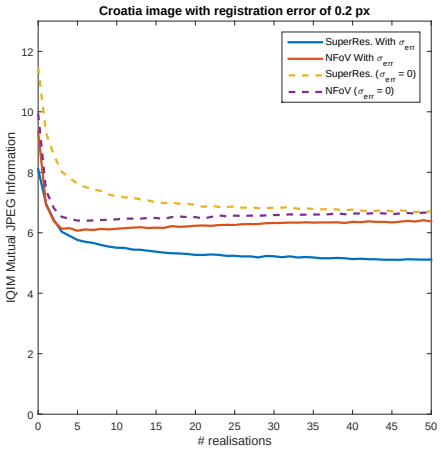
Figure C.10: Information quality information metric JPEG (IQIM JPEG) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.10(a) - C.10(k))


Croatia image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

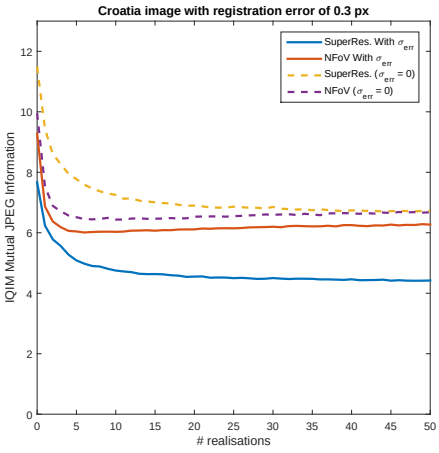
(a)


Croatia image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

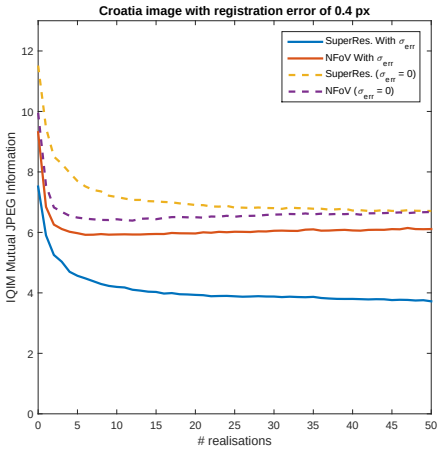
(b)


Croatia image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

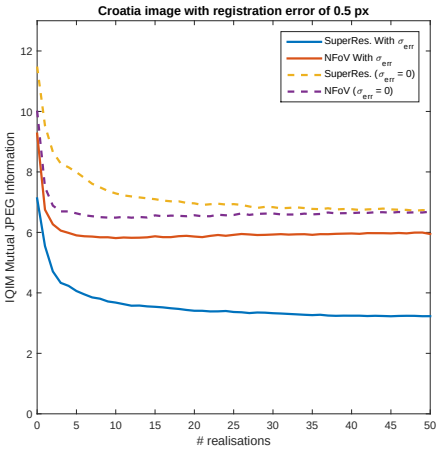
(c)


Croatia image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

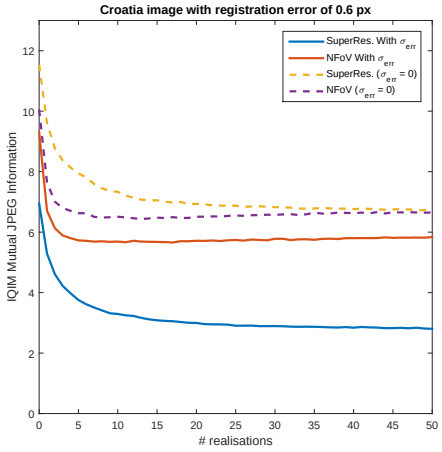
(d)


Croatia image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

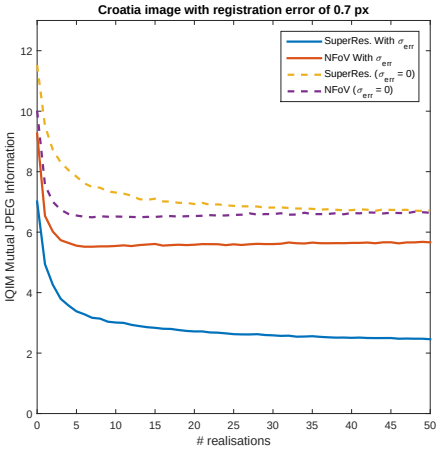
(e)


Croatia image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

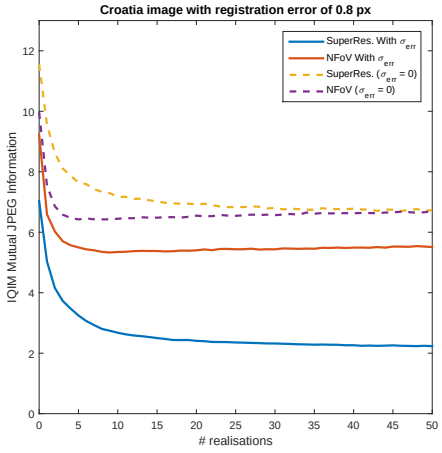
(f)



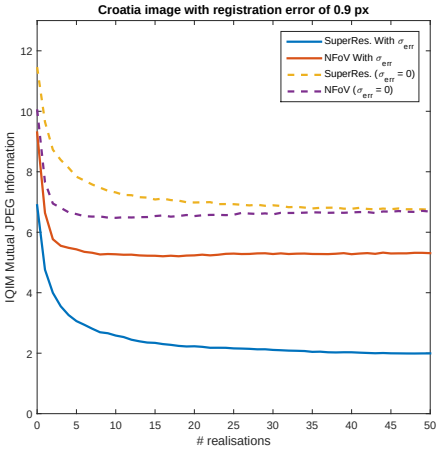
(g)



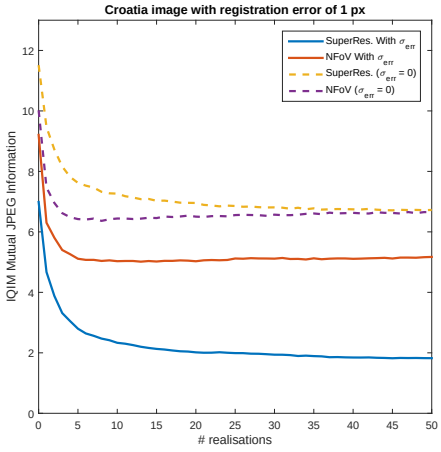
(h)



(i)

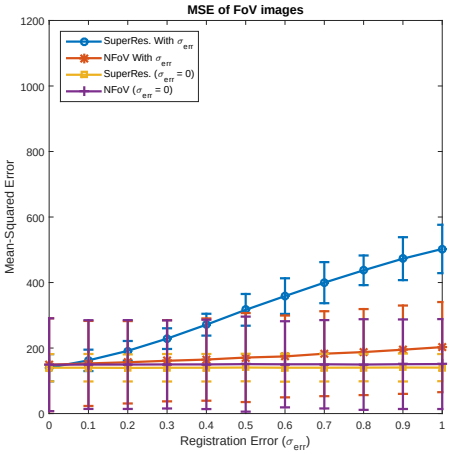


(j)

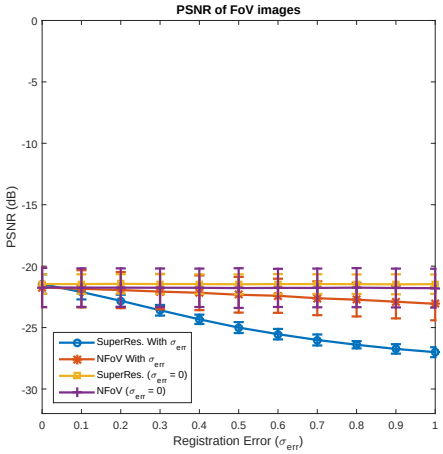


(k)

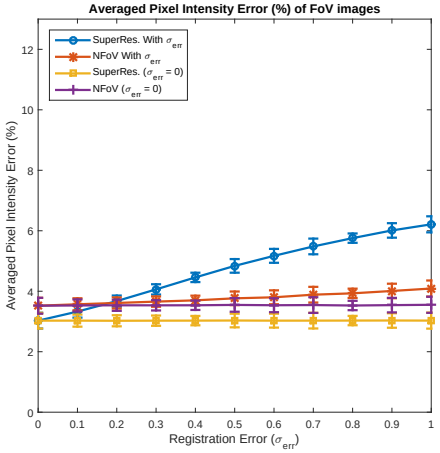
Figure C.11: Information quality information metric (mutual) JPEG (IQIMM JPEG) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.11(a) - C.11(k))


MSE of FoV images
Mean-Squared Error
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

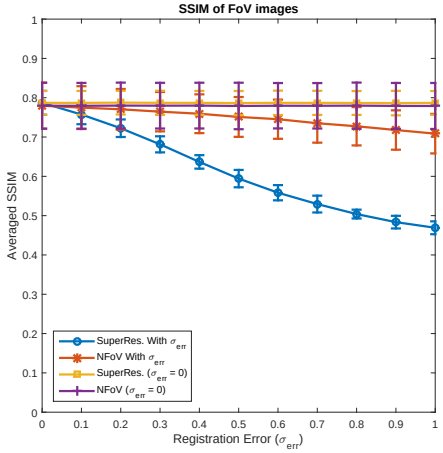
(a)


PSNR of FoV images
PSNR (dB)
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

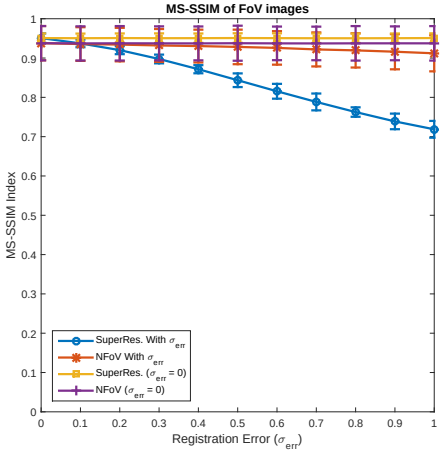
(b)


Averaged Pixel Intensity Error of FoV images
Averaged Pixel Intensity Error (%)
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

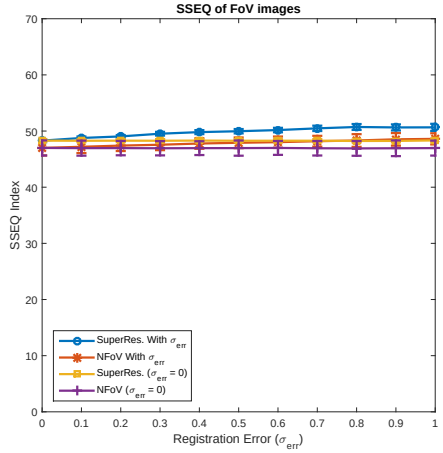
(c)


SSIM of FoV images
Averaged SSIM
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

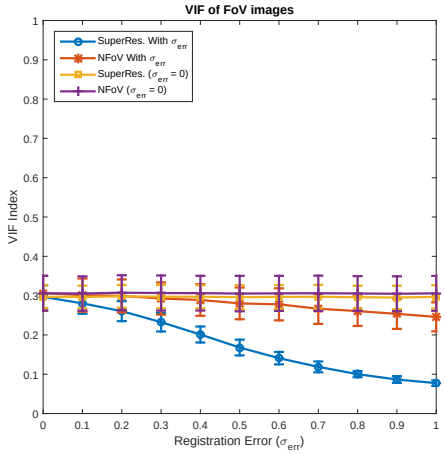
(d)


MS-SSIM of FoV images
MS-SSIM Index
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

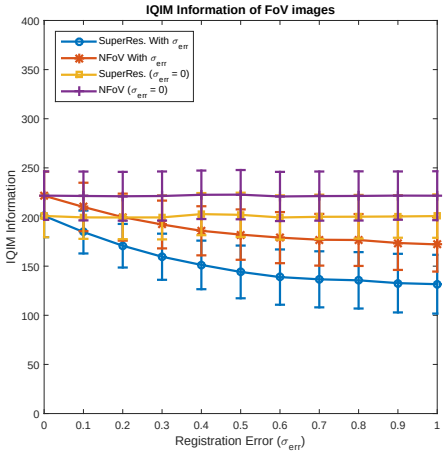
(e)


SSEQ of FoV images
SSEQ Index
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

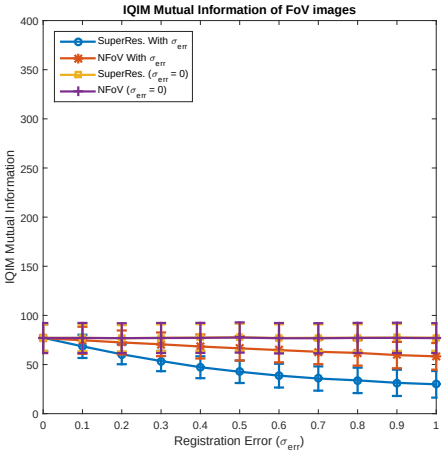
(f)



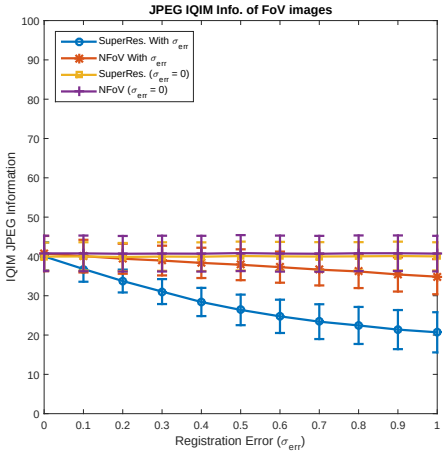
(g)



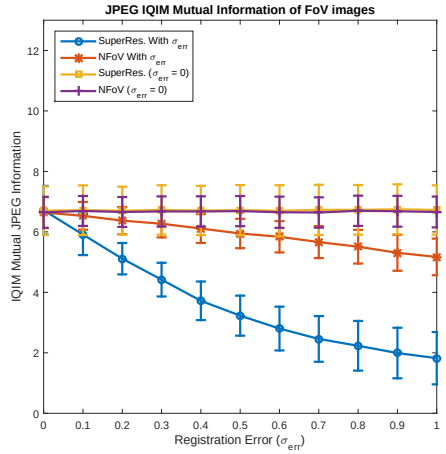
(h)



(i)

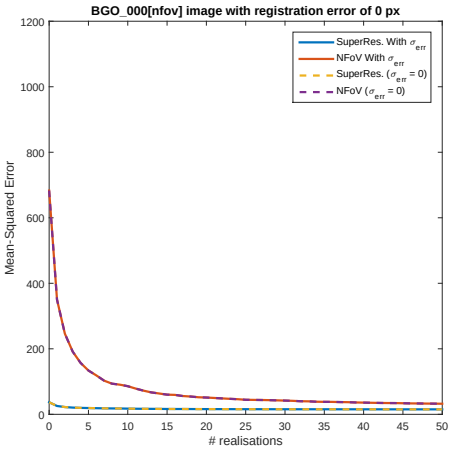


(j)

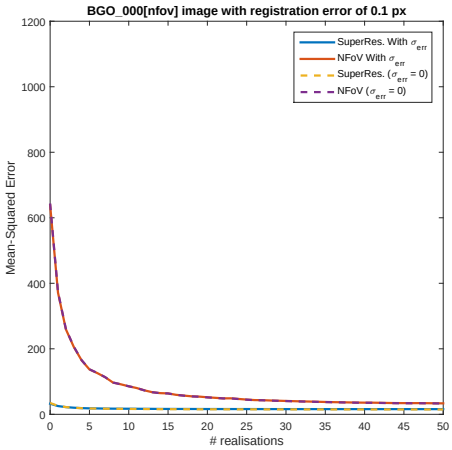


(k)

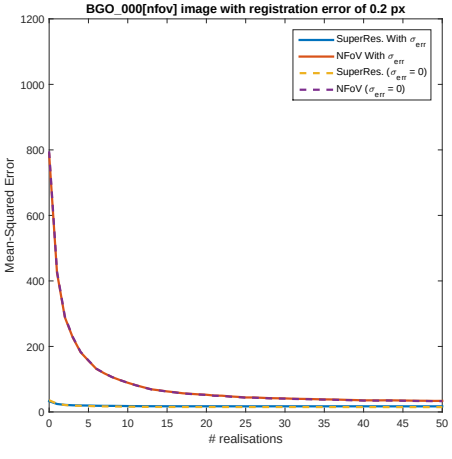
Figure C.12: Figures C.12(a) - C.12(k) showing the different image quality metrics used with error bars for the Croatia image. Each sub-figure contains the data for the perfect registration case, and the case with a registration error of $0.5px$, for the TA and SR techniques


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

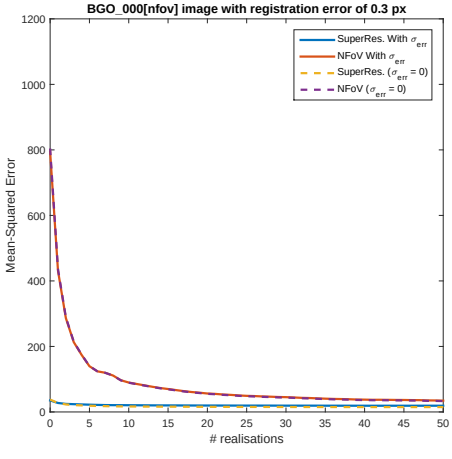
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

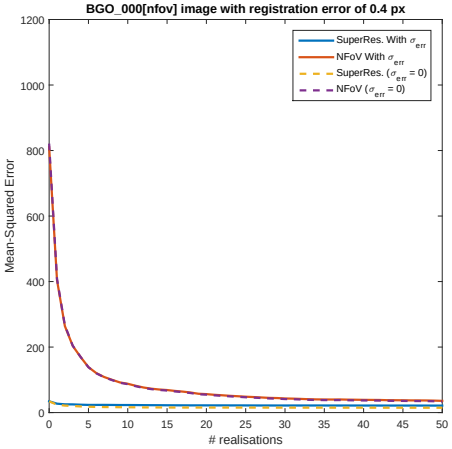
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

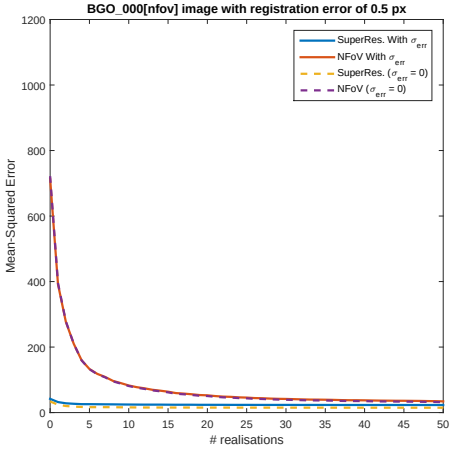
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

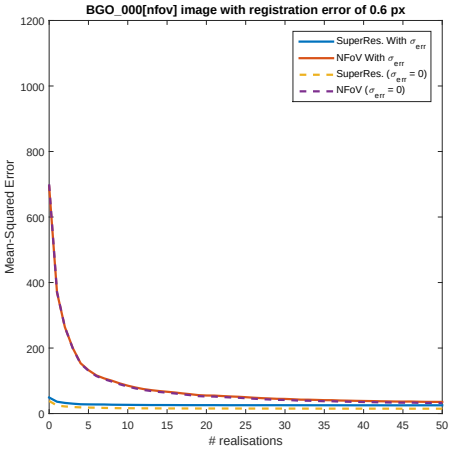
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

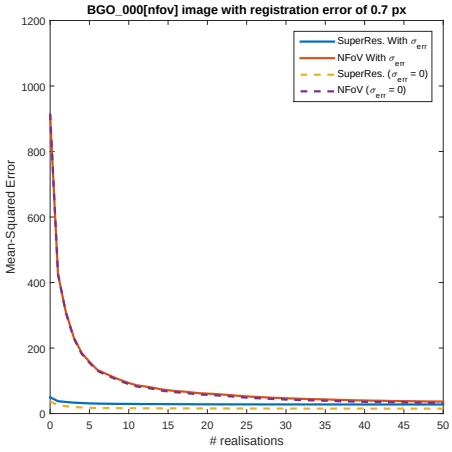
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Mean-Squared Error
realisations

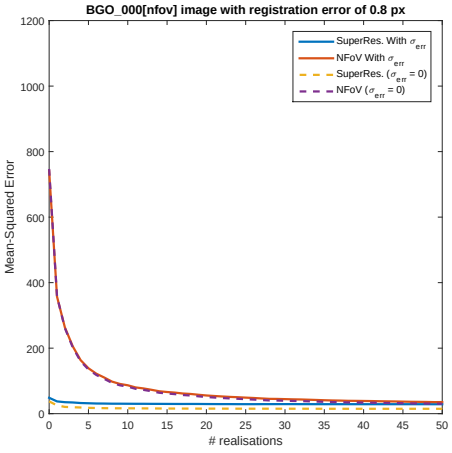
(f)



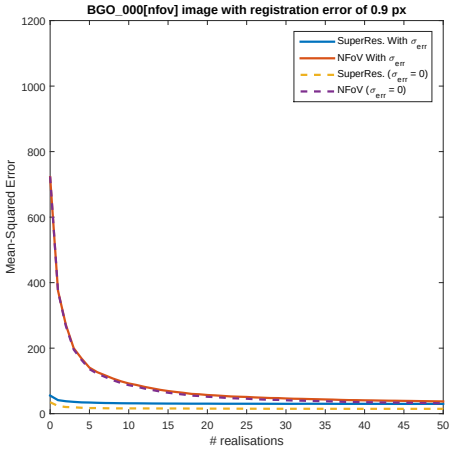
(g)



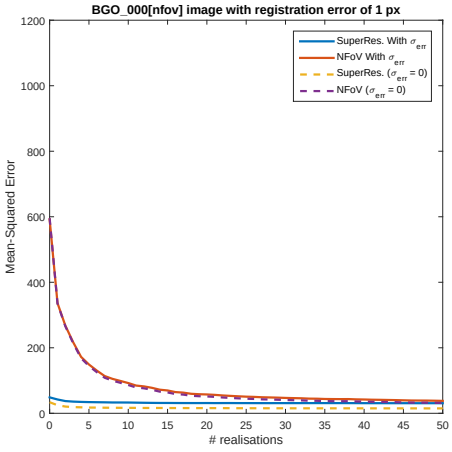
(h)



(i)

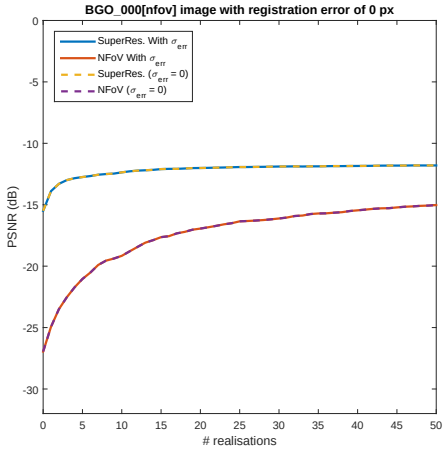


(j)

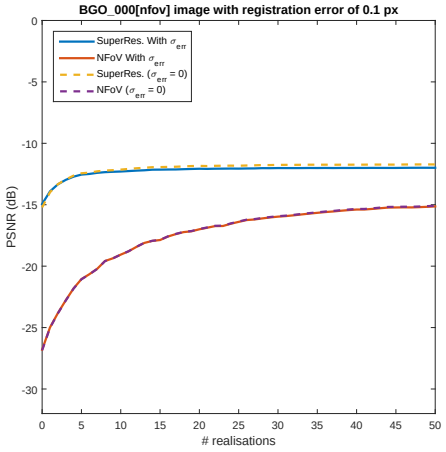


(k)

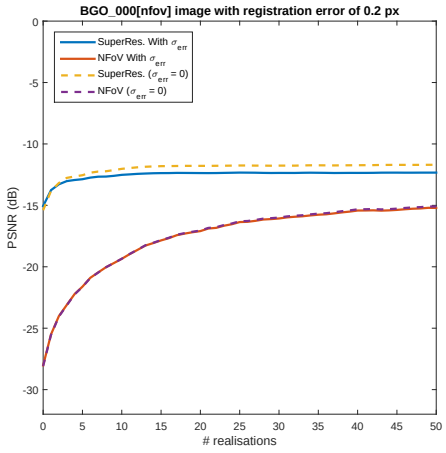
Figure C.13: Mean-squared error (MSE) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.13(a) - C.13(k))


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

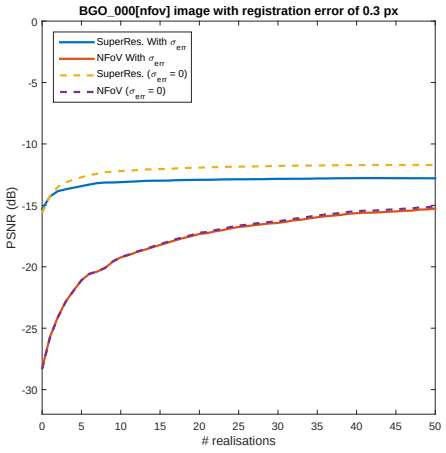
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

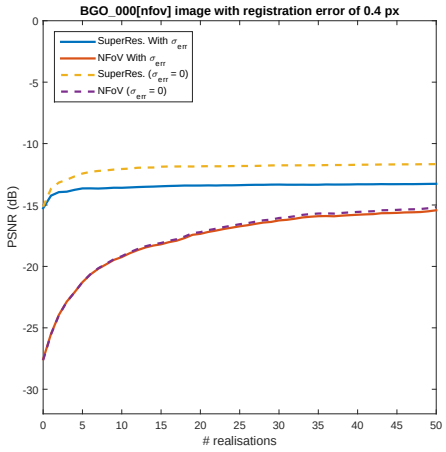
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

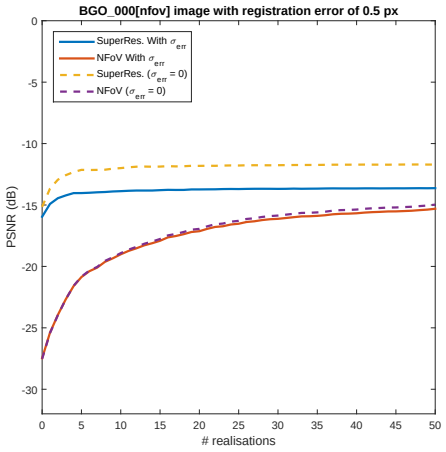
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

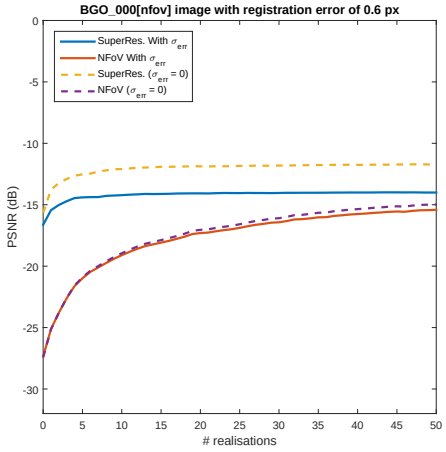
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

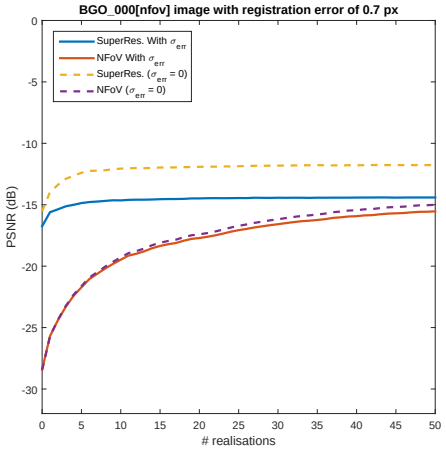
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
PSNR (dB)
realisations

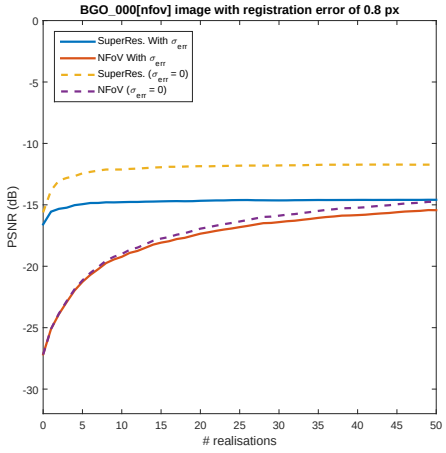
(f)



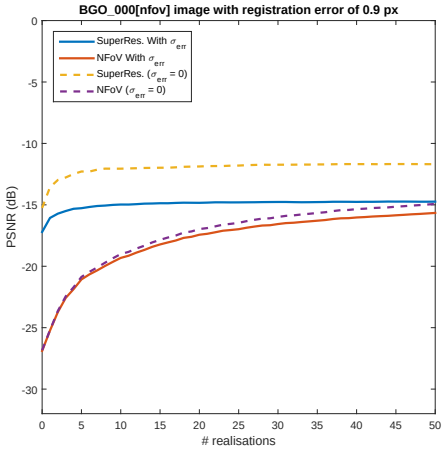
(g)



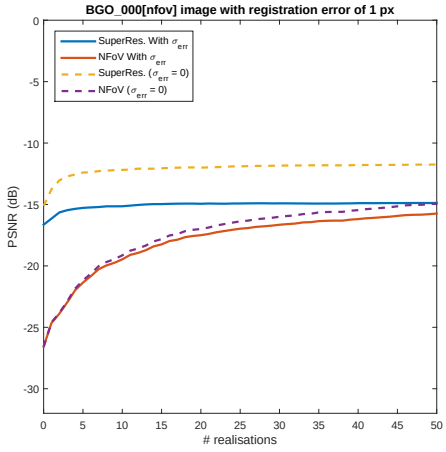
(h)



(i)

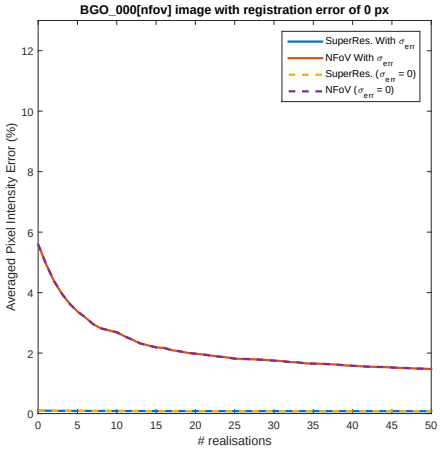


(j)

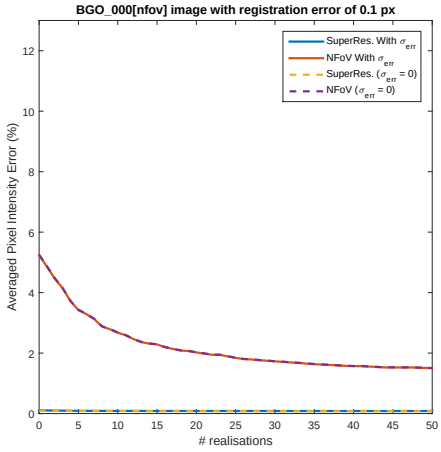


(k)

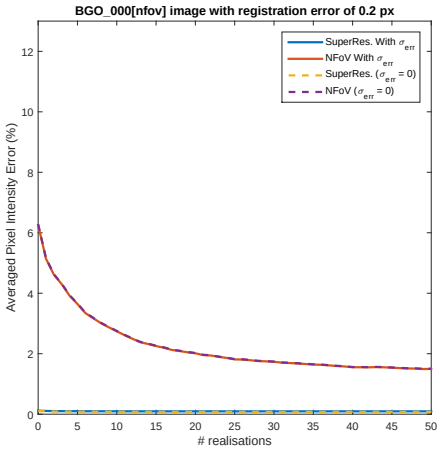
Figure C.14: Peak signal-to-noise ratio (PSNR) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.14(a) - C.14(k))


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

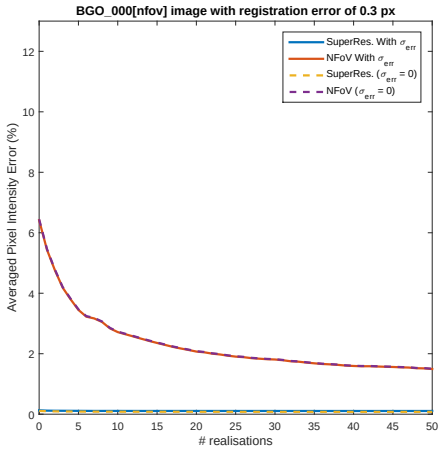
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

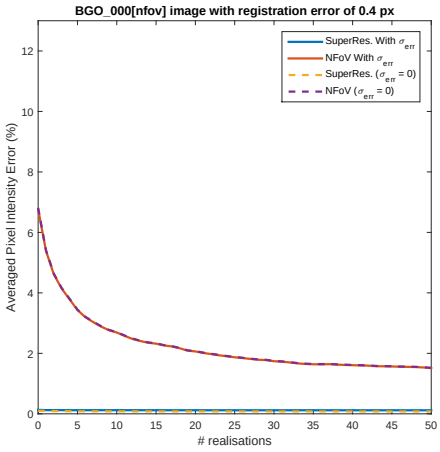
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

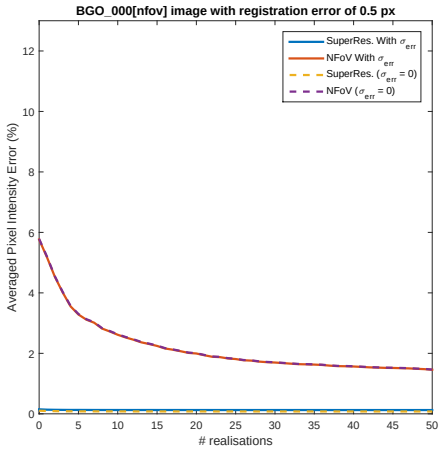
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

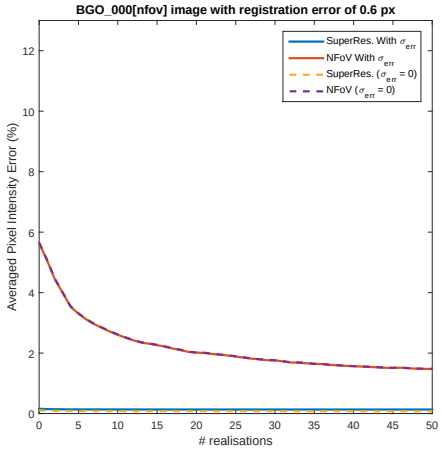
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

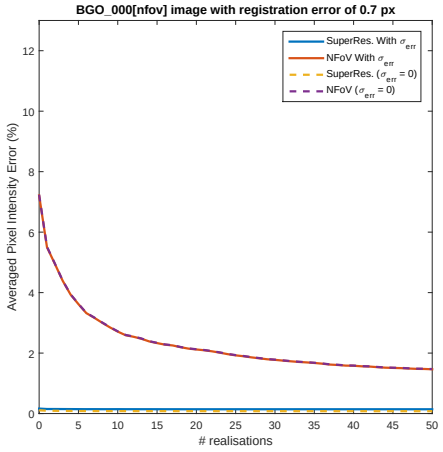
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
Averaged Pixel Intensity Error (%)
realisations

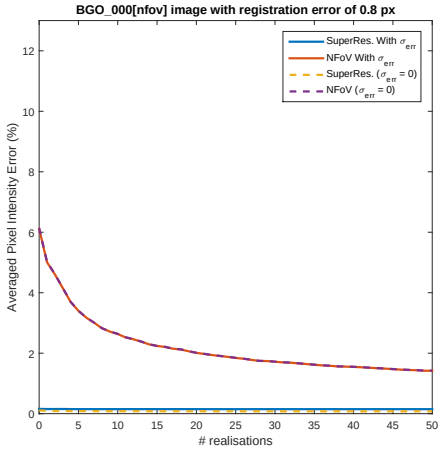
(f)



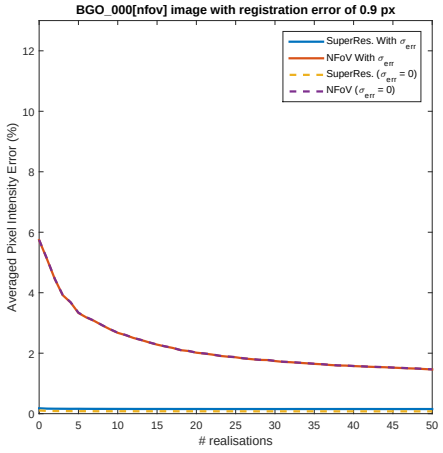
(g)



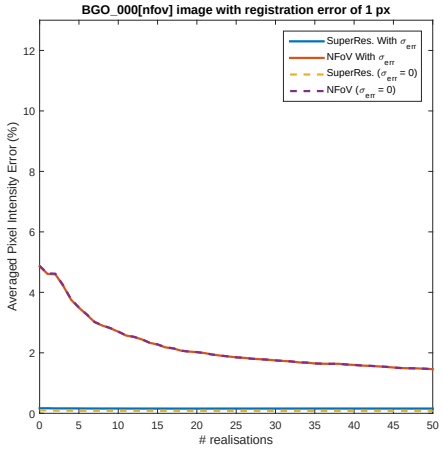
(h)



(i)

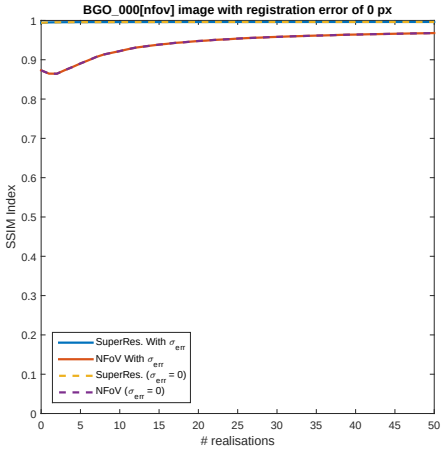


(j)

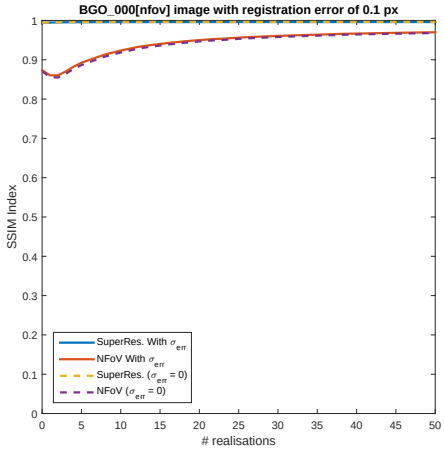


(k)

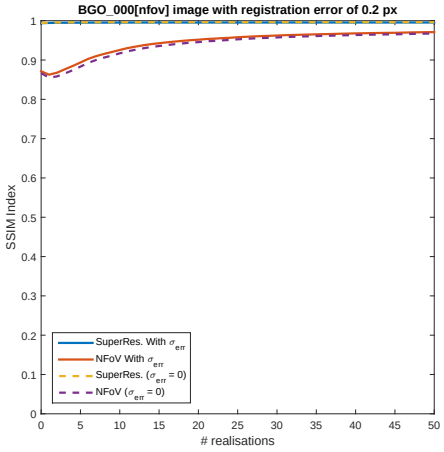
Figure C.15: Averaged pixel intensity error against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.15(a) - C.15(k))


BGO_000[nfov] image with registration error of 0 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

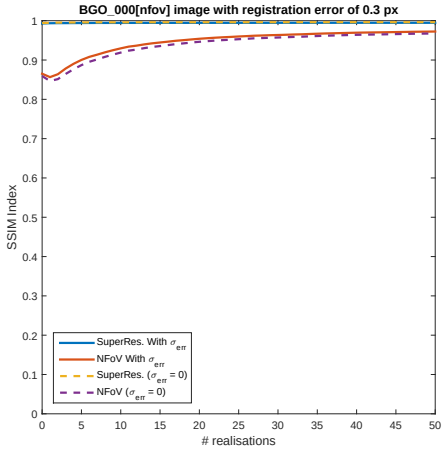
(a)


BGO_000[nfov] image with registration error of 0.1 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

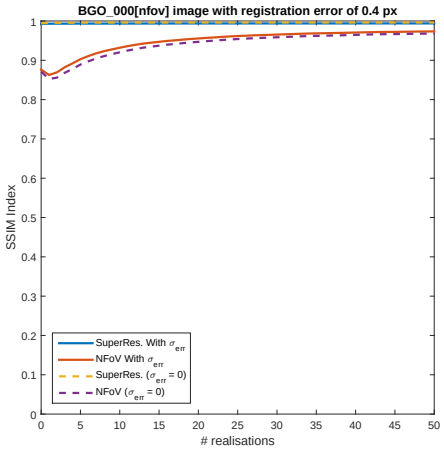
(b)


BGO_000[nfov] image with registration error of 0.2 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

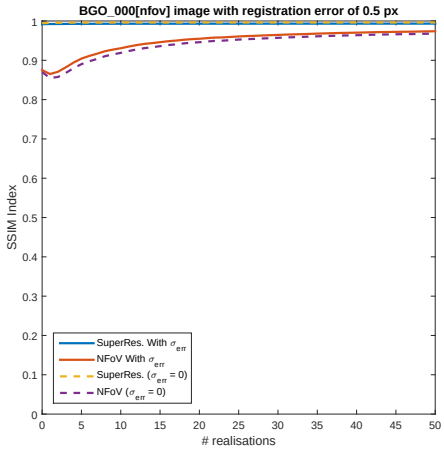
(c)


BGO_000[nfov] image with registration error of 0.3 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

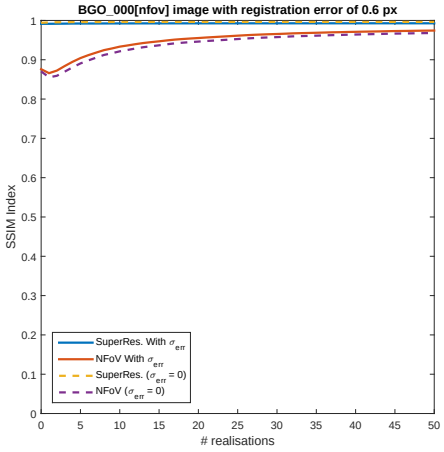
(d)


BGO_000[nfov] image with registration error of 0.4 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

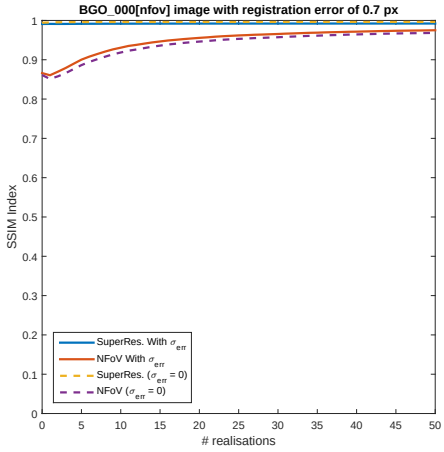
(e)


BGO_000[nfov] image with registration error of 0.5 px
SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

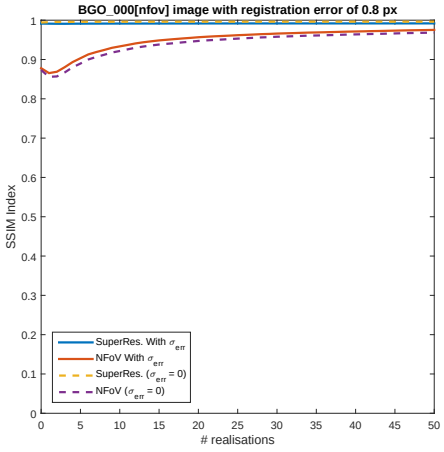
(f)



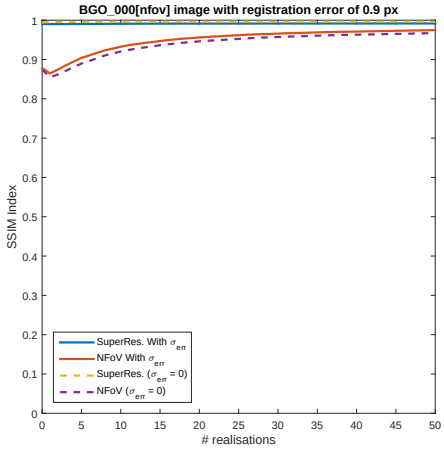
(g)



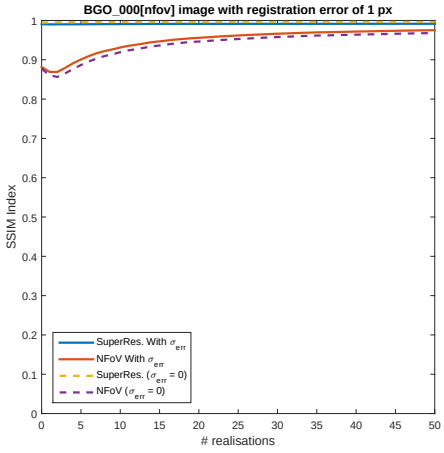
(h)



(i)

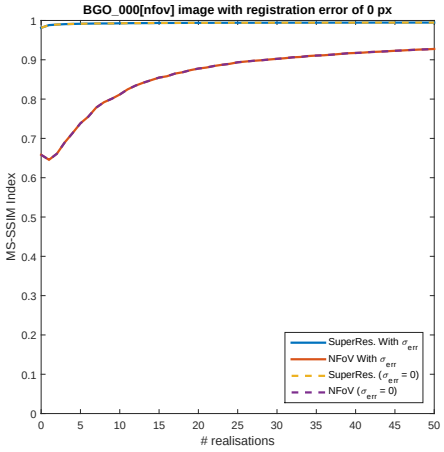


(j)

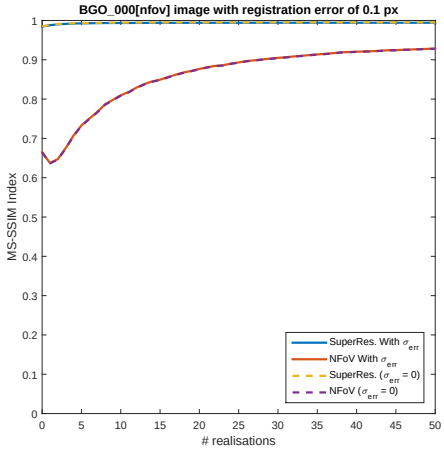


(k)

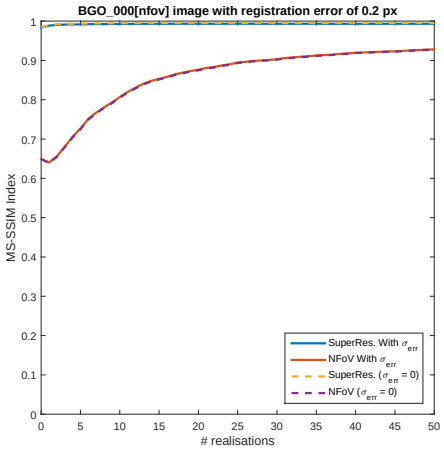
Figure C.16: Structural similarity (SSIM) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.16(a) - C.16(k))


BGO_000[nfov] image with registration error of 0 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

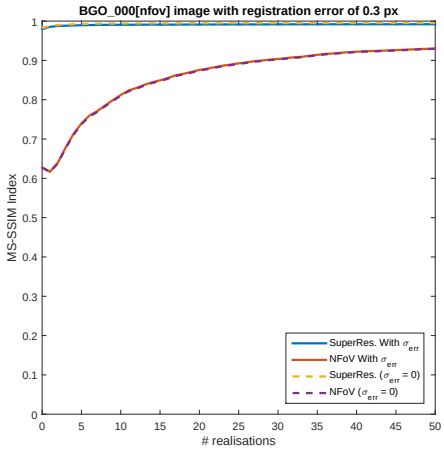
(a)


BGO_000[nfov] image with registration error of 0.1 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

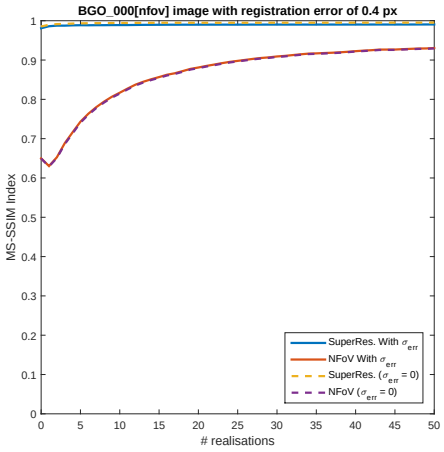
(b)


BGO_000[nfov] image with registration error of 0.2 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

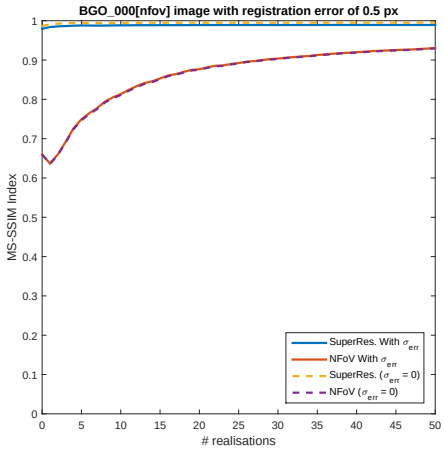
(c)


BGO_000[nfov] image with registration error of 0.3 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

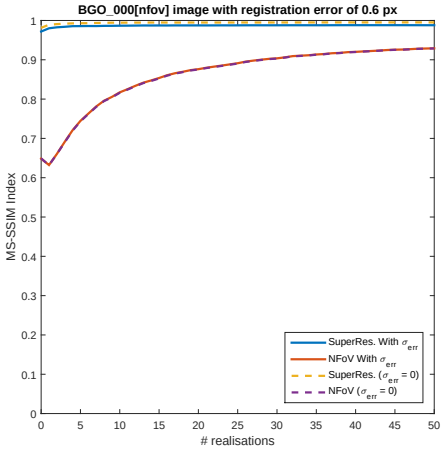
(d)


BGO_000[nfov] image with registration error of 0.4 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

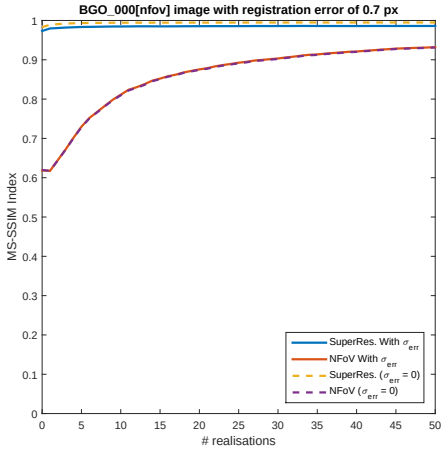
(e)


BGO_000[nfov] image with registration error of 0.5 px
MS-SSIM Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

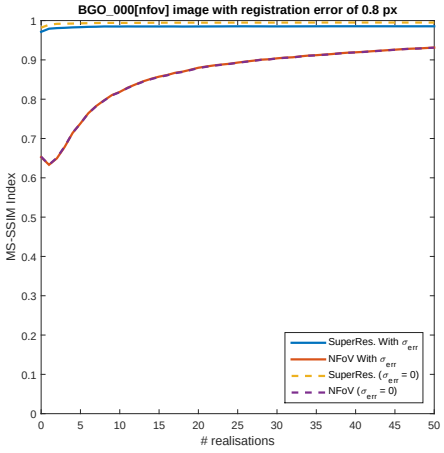
(f)



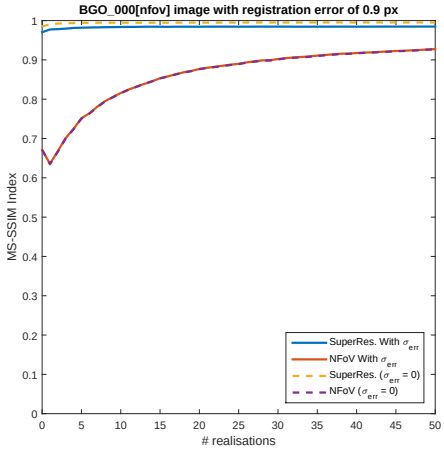
(g)



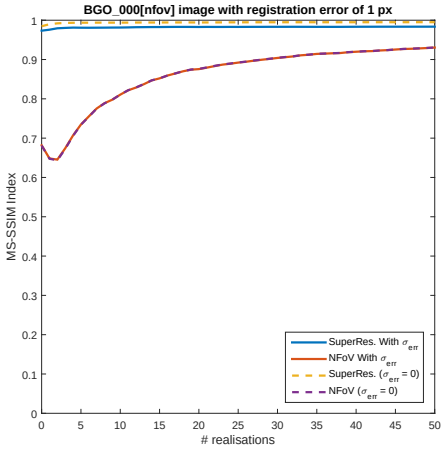
(h)



(i)

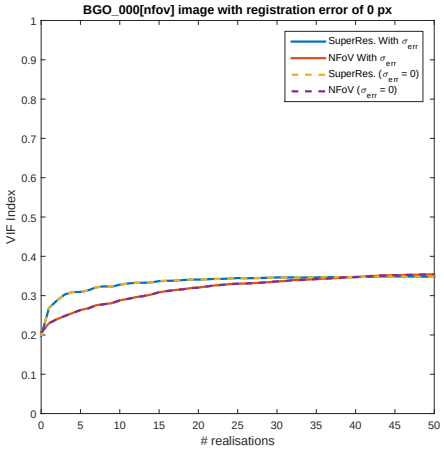


(j)

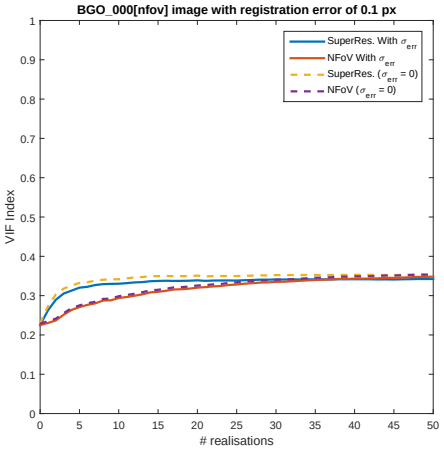


(k)

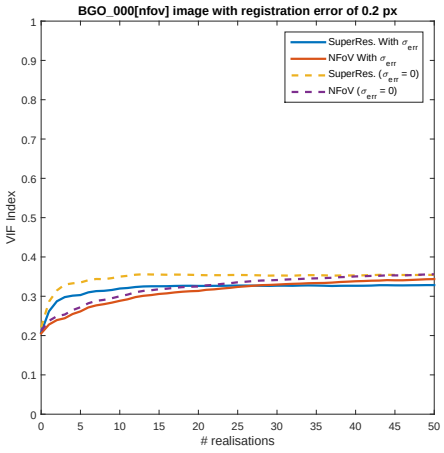
Figure C.17: Multi-scale structural similarity (MS-SSIM) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.17(a) - C.17(k))


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

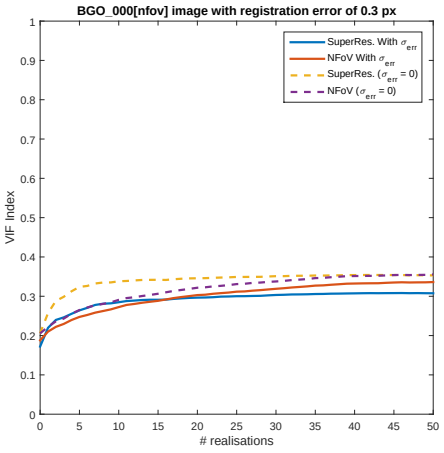
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

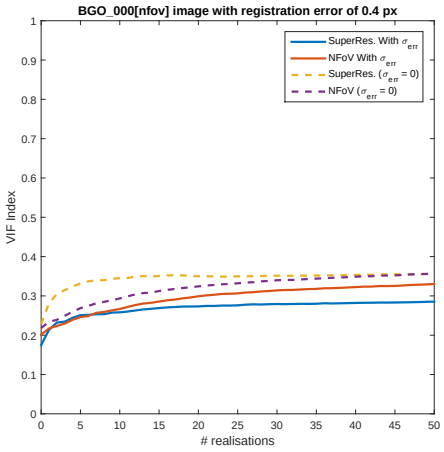
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

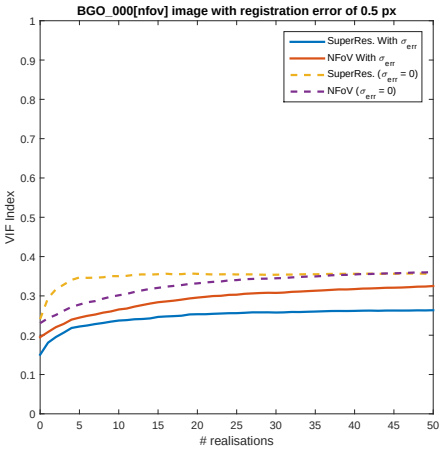
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

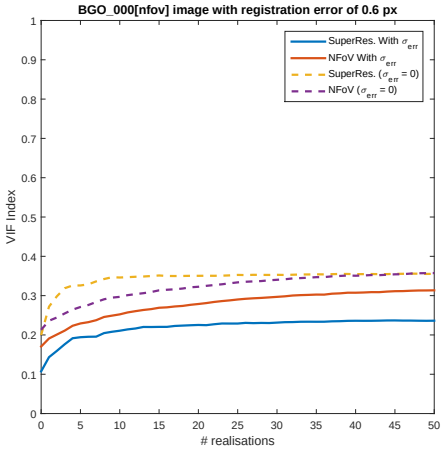
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

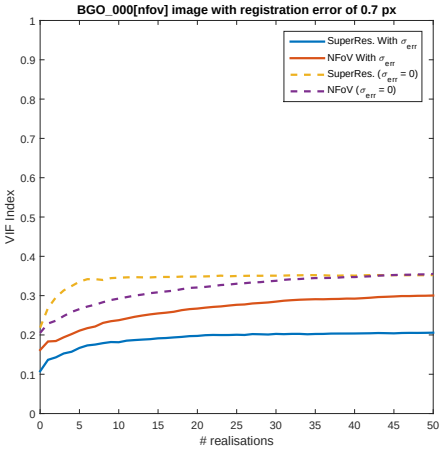
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
VIF Index
realisations

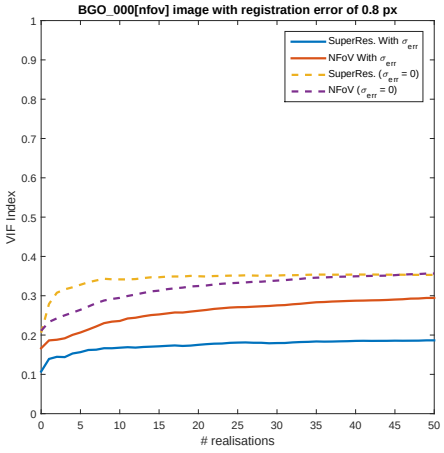
(f)



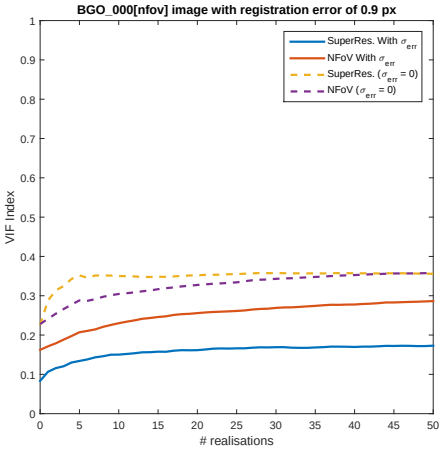
(g)



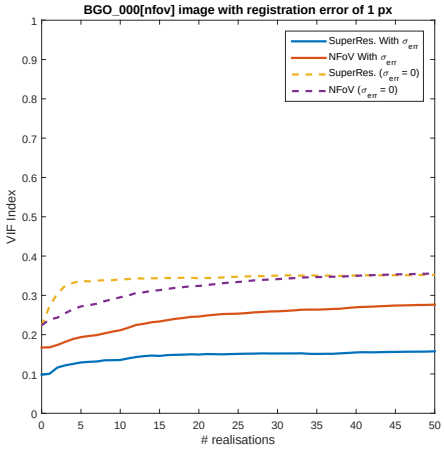
(h)



(i)

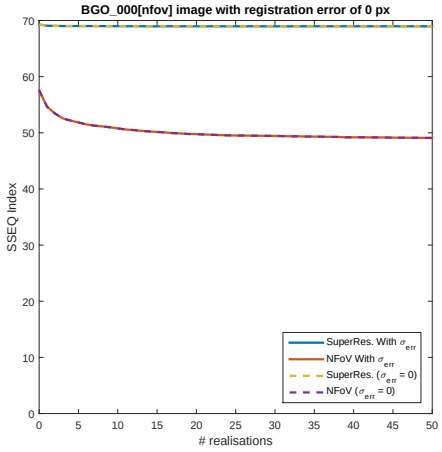


(j)

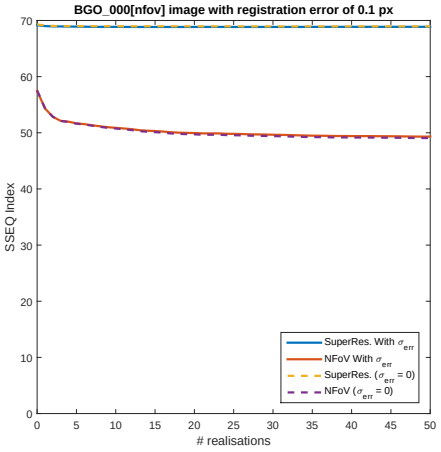


(k)

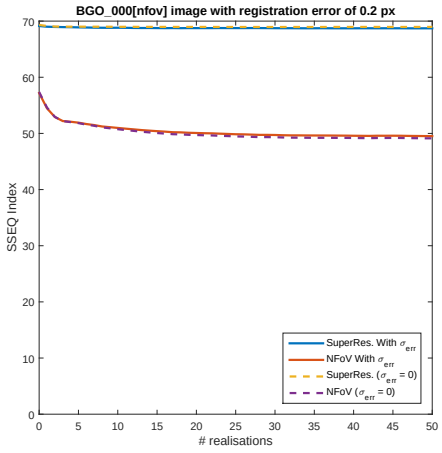
Figure C.18: Visual information fidelity (VIF) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.18(a) - C.18(k))


BGO_000[nfov] image with registration error of 0 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

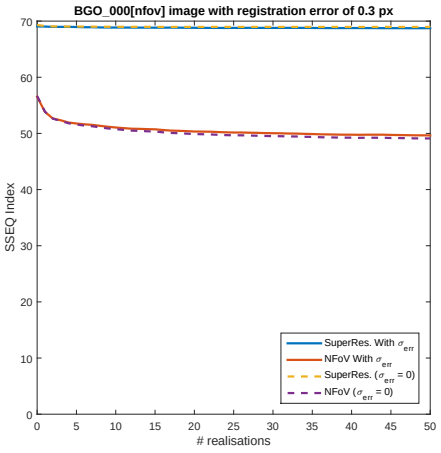
(a)


BGO_000[nfov] image with registration error of 0.1 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

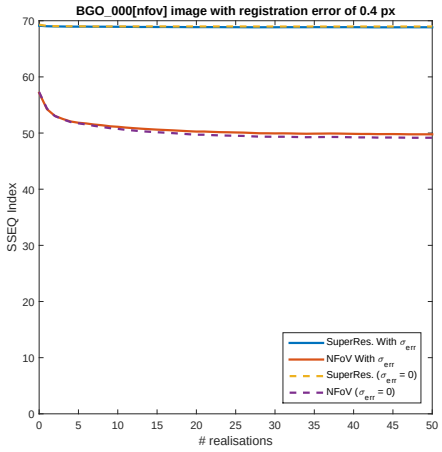
(b)


BGO_000[nfov] image with registration error of 0.2 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

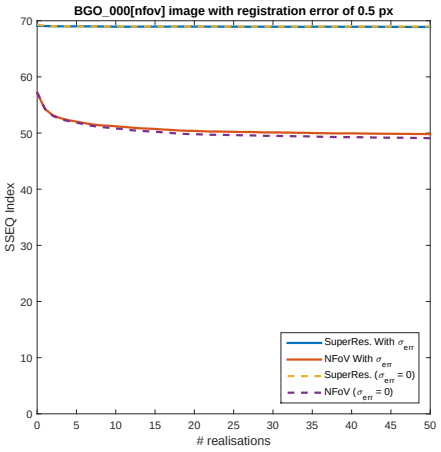
(c)


BGO_000[nfov] image with registration error of 0.3 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

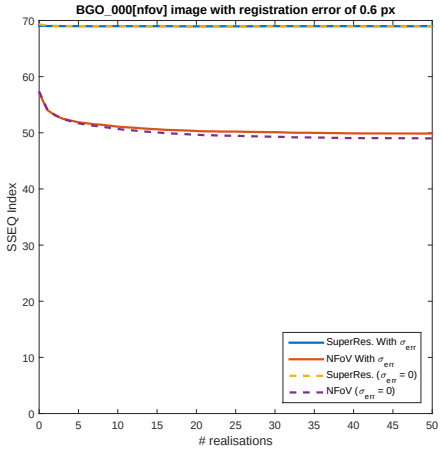
(d)


BGO_000[nfov] image with registration error of 0.4 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

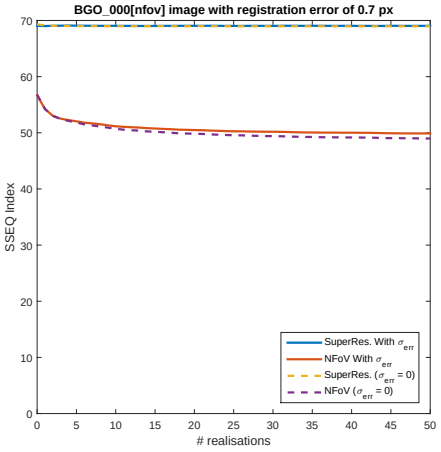
(e)


BGO_000[nfov] image with registration error of 0.5 px
SSEQ Index
realisations
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

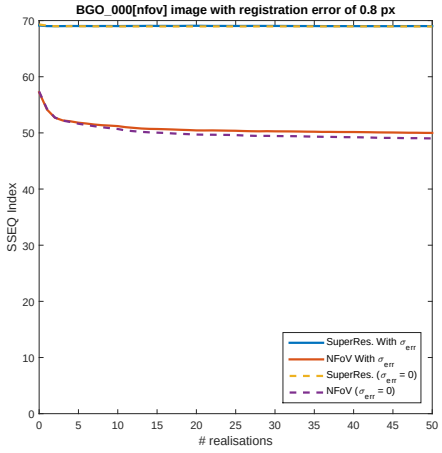
(f)



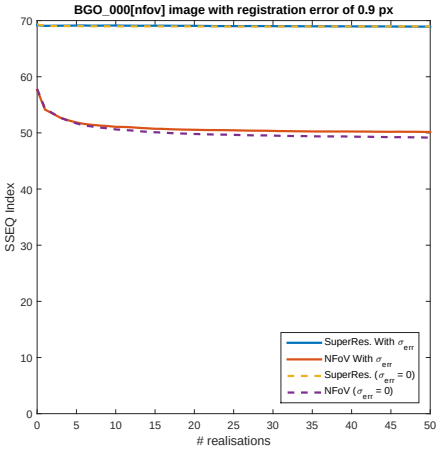
(g)



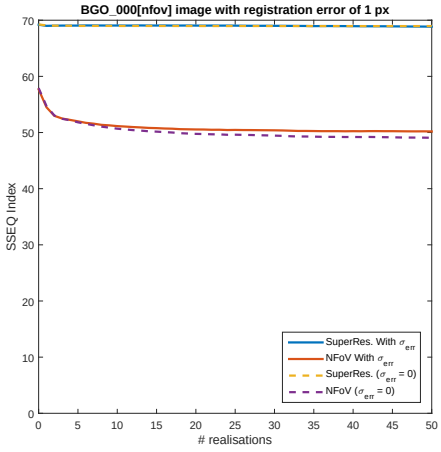
(h)



(i)

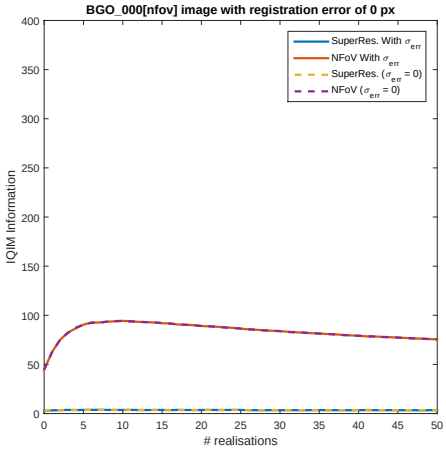


(j)

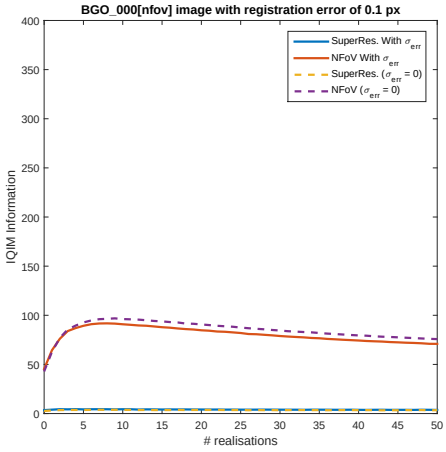


(k)

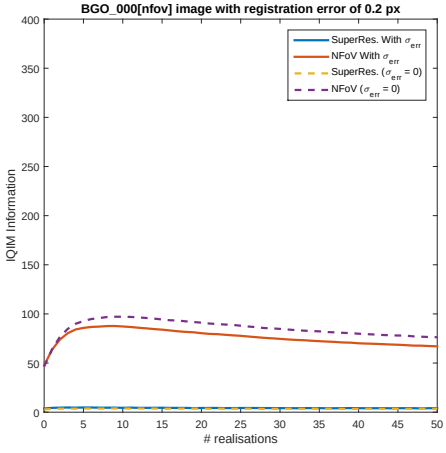
Figure C.19: Spatial-spectral entropy-based quality (SSEQ) index against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.19(a) - C.19(k))


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQM Information
realisations

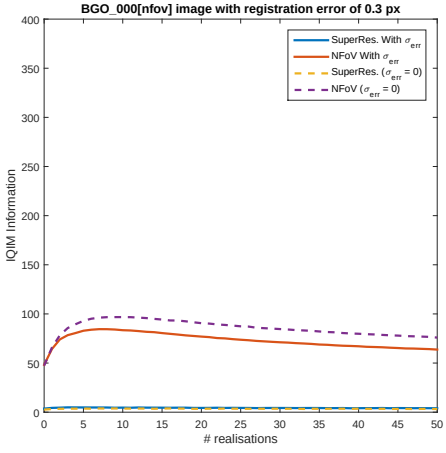
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQM Information
realisations

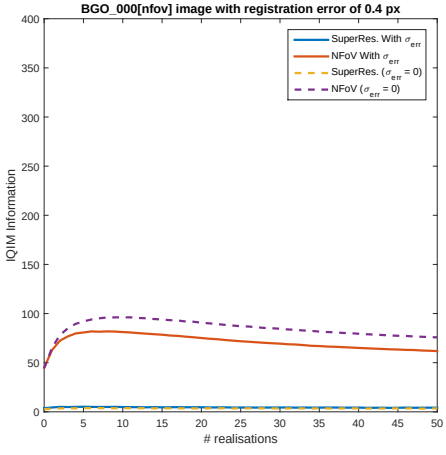
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQM Information
realisations

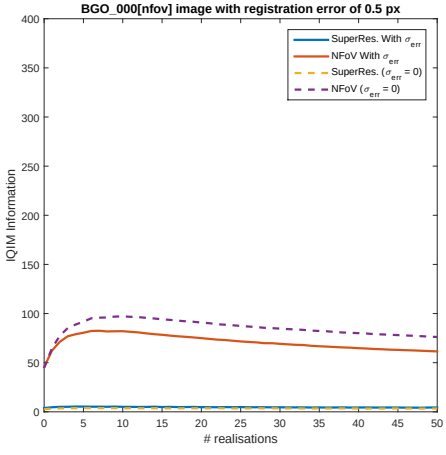
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQM Information
realisations

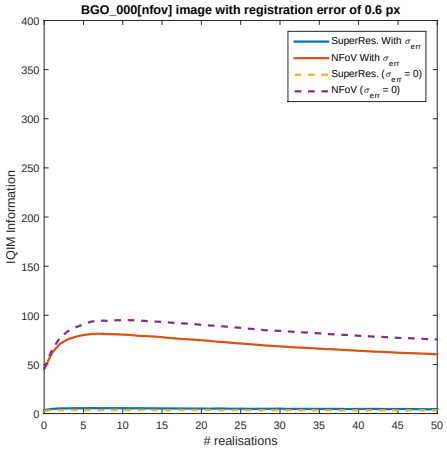
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQM Information
realisations

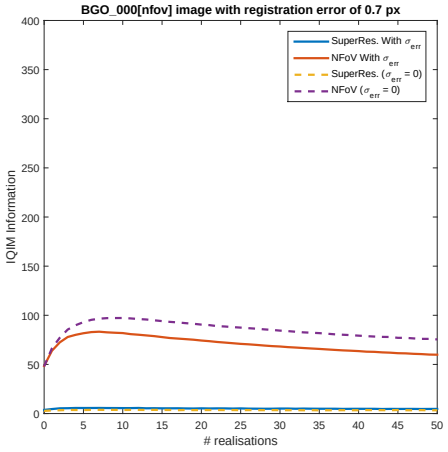
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQM Information
realisations

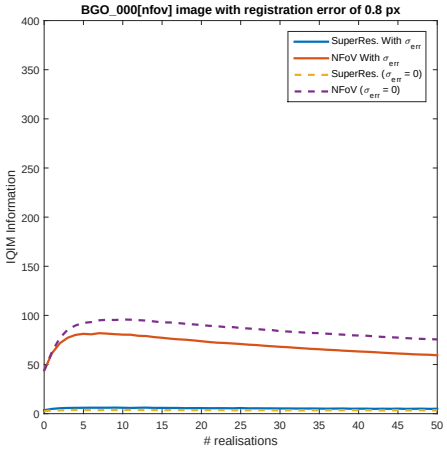
(f)



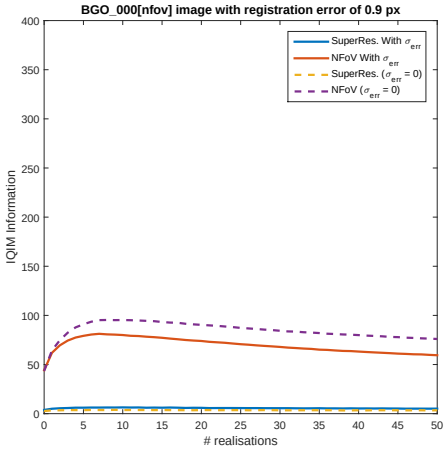
(g)



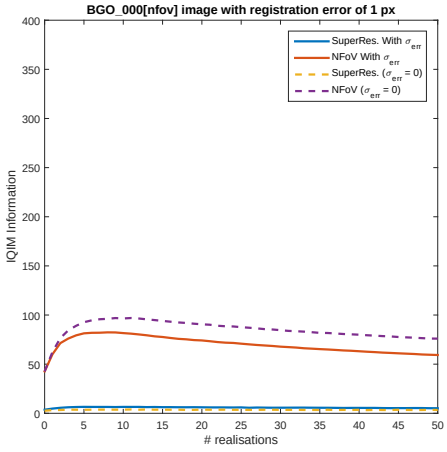
(h)



(i)

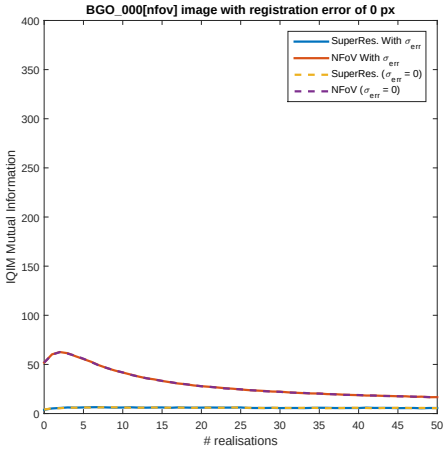


(j)

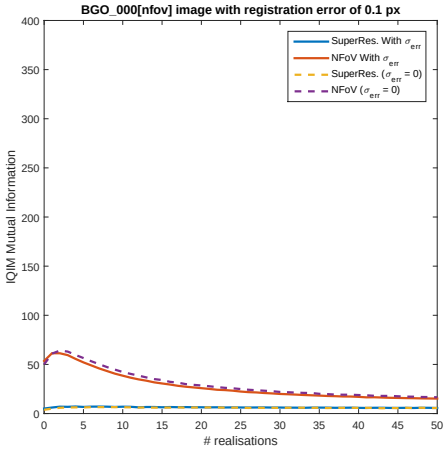


(k)

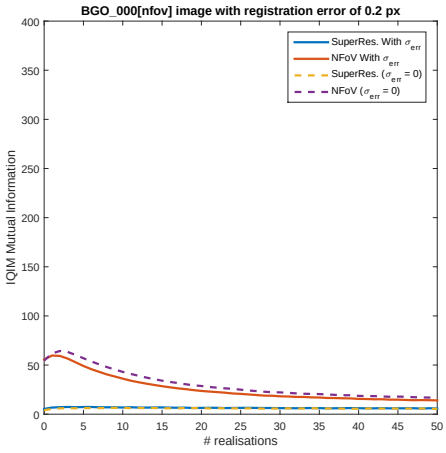
Figure C.20: Information quality information metric (IQIM) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.20(a) - C.20(k))


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

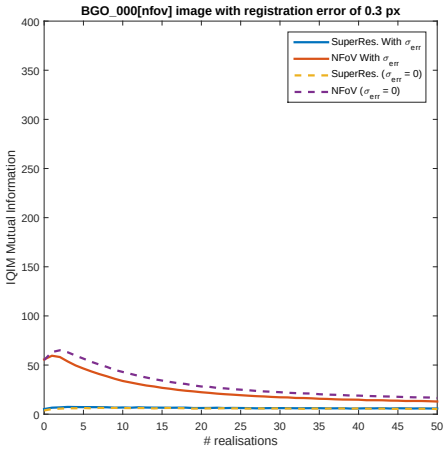
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

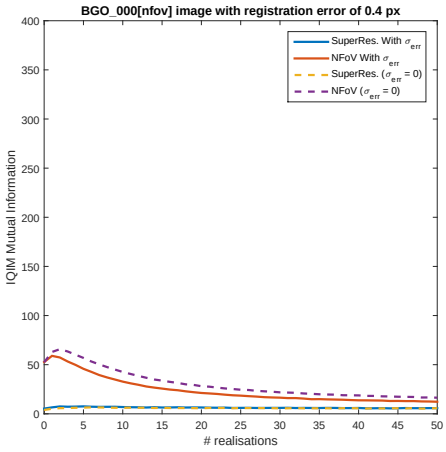
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

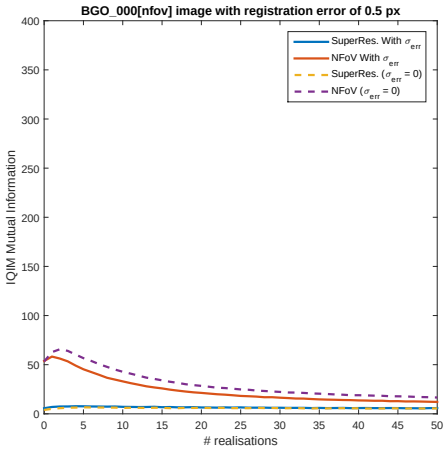
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

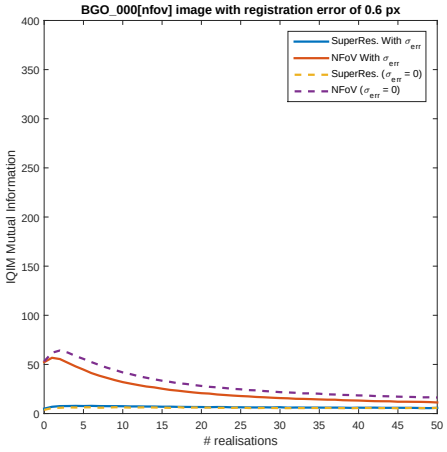
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

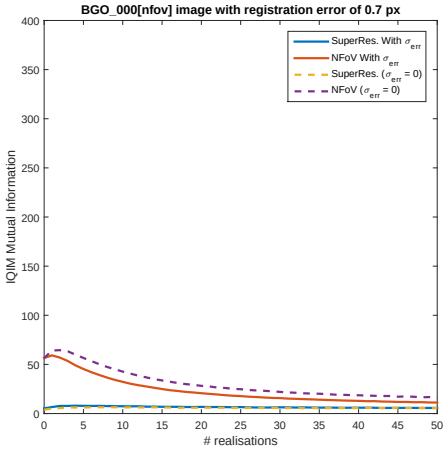
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual Information
realisations

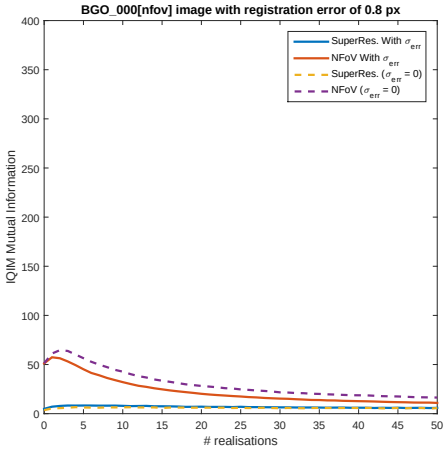
(f)



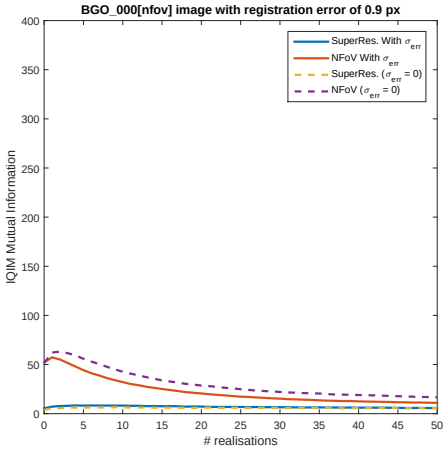
(g)



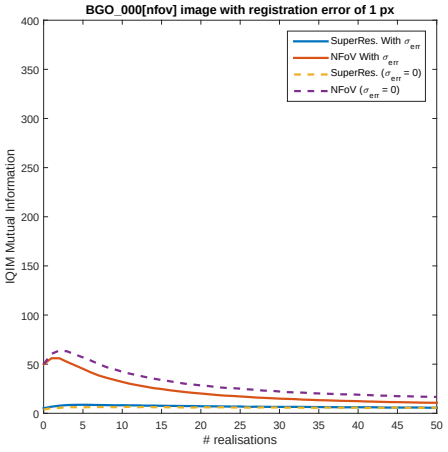
(h)



(i)

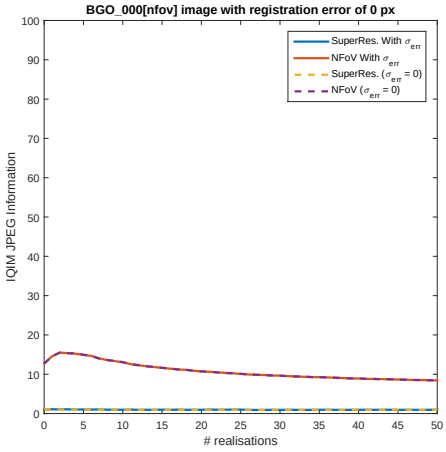


(j)

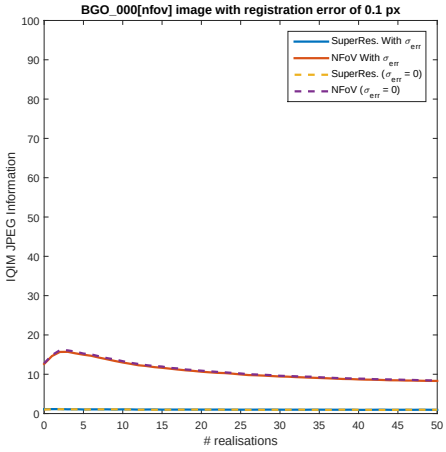


(k)

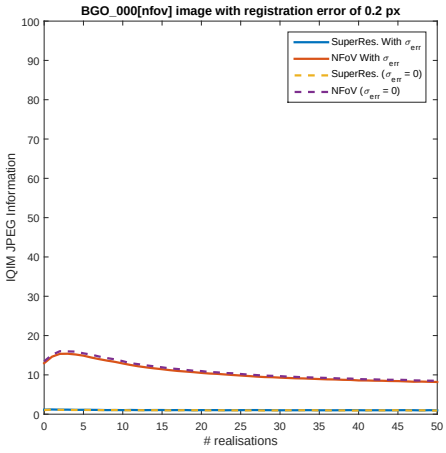
Figure C.21: Information quality information metric (mutual) (IQIMM) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.21(a) - C.21(k))


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

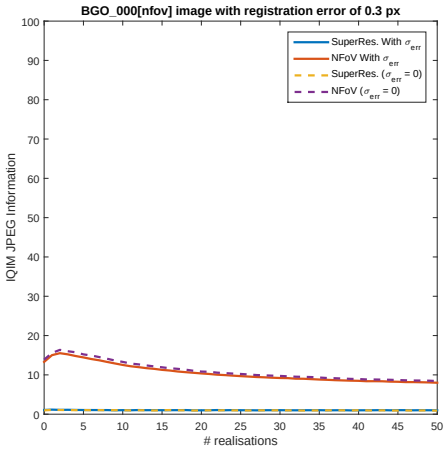
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

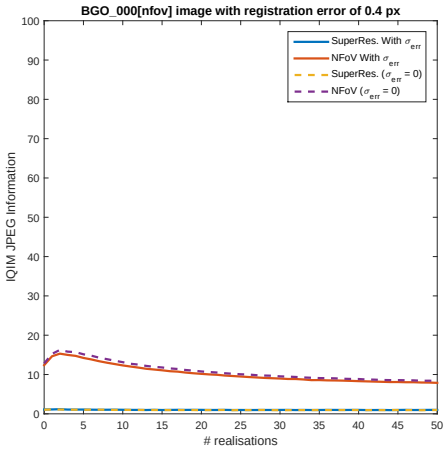
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

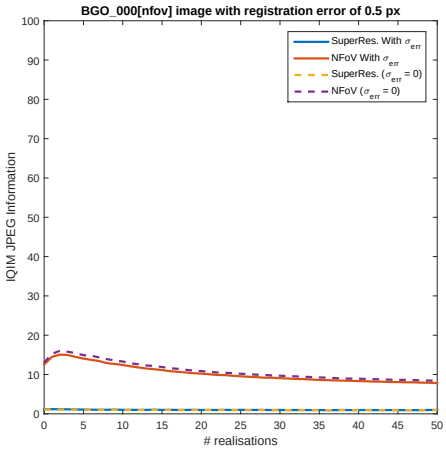
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

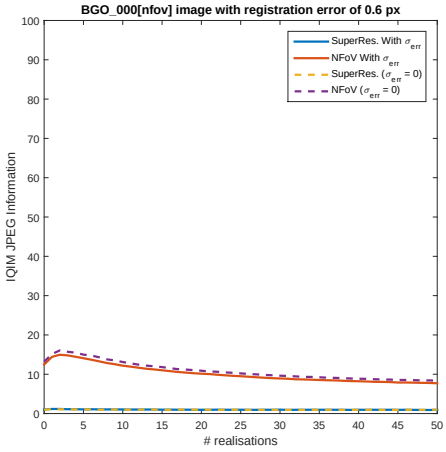
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

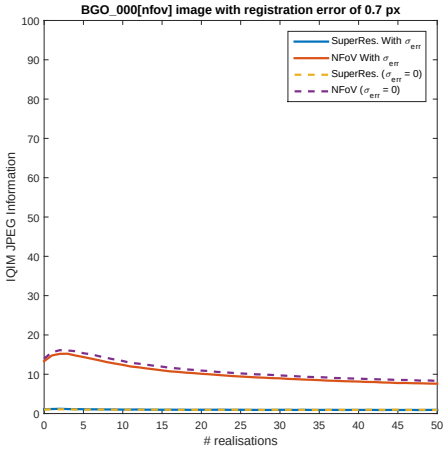
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM JPEG Information
realisations

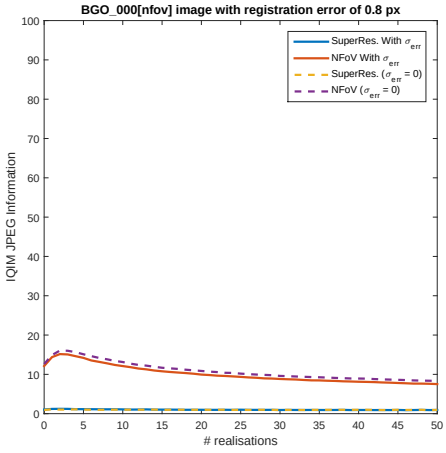
(f)



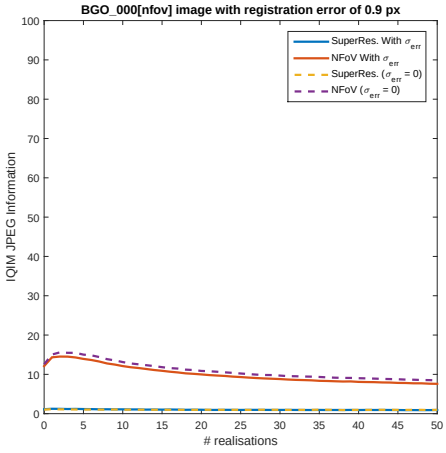
(g)



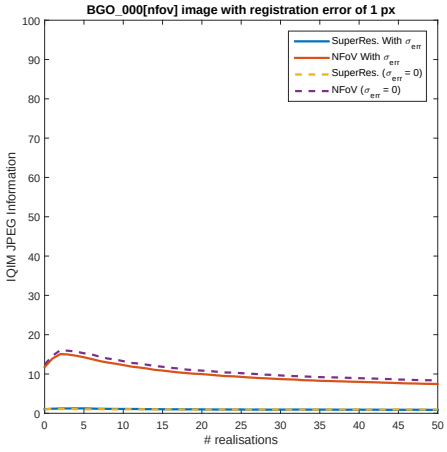
(h)



(i)

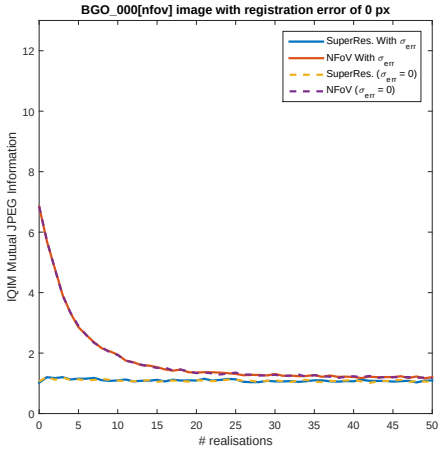


(j)

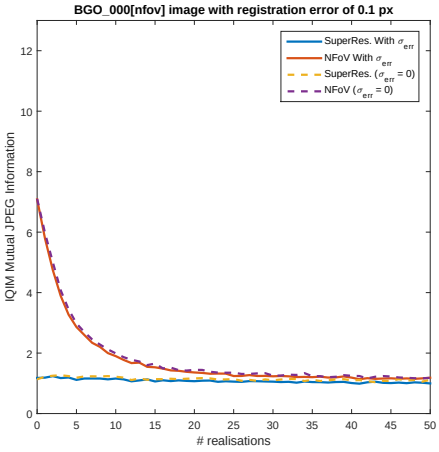


(k)

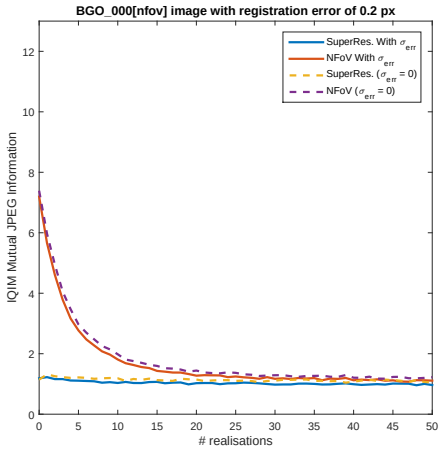
Figure C.22: Information quality information metric JPEG (IQIM JPEG) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.22(a) - C.22(k))


BGO_000[nfov] image with registration error of 0 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

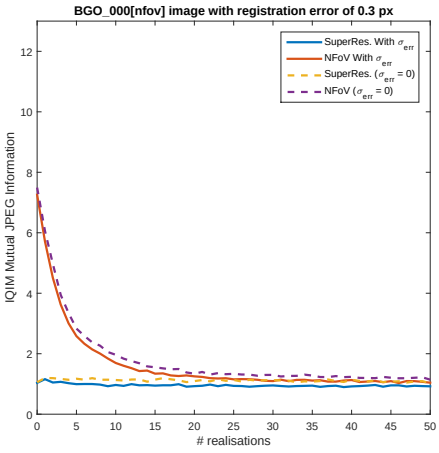
(a)


BGO_000[nfov] image with registration error of 0.1 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

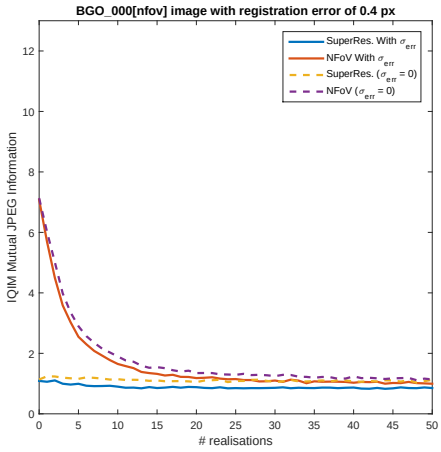
(b)


BGO_000[nfov] image with registration error of 0.2 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

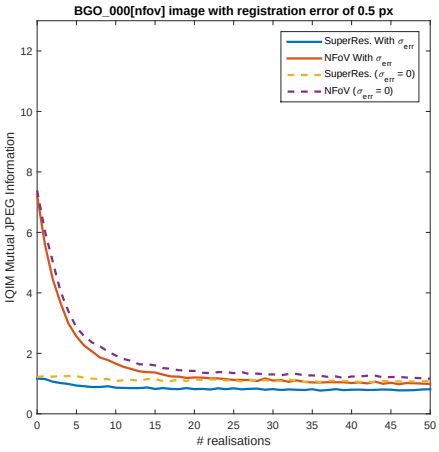
(c)


BGO_000[nfov] image with registration error of 0.3 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

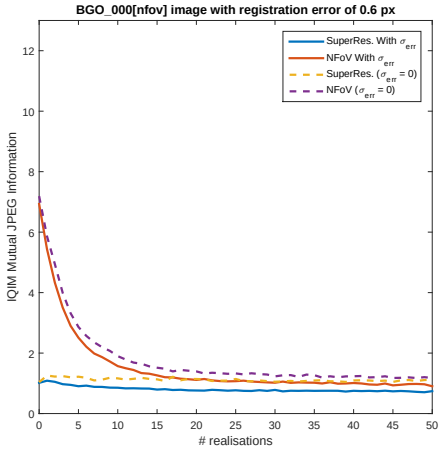
(d)


BGO_000[nfov] image with registration error of 0.4 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

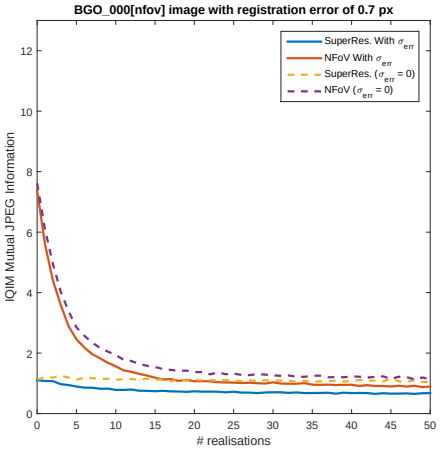
(e)


BGO_000[nfov] image with registration error of 0.5 px
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)
IQIM Mutual JPEG Information
realisations

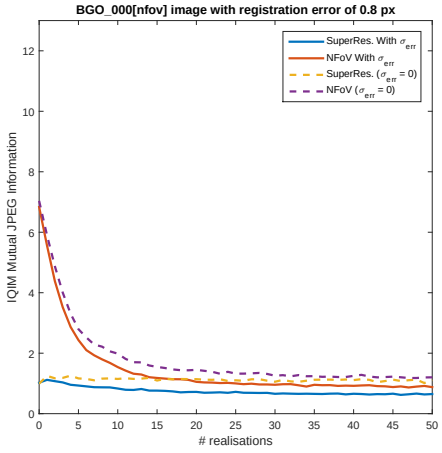
(f)



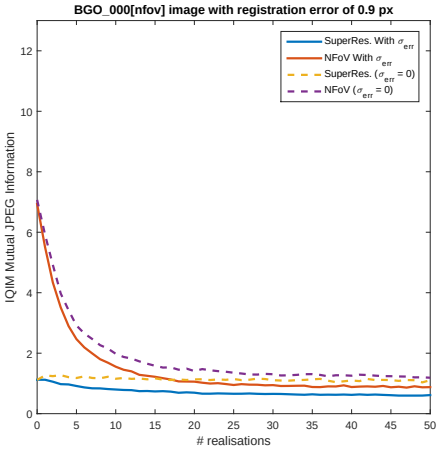
(g)



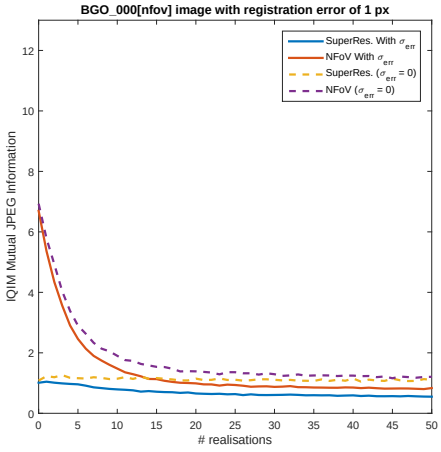
(h)



(i)

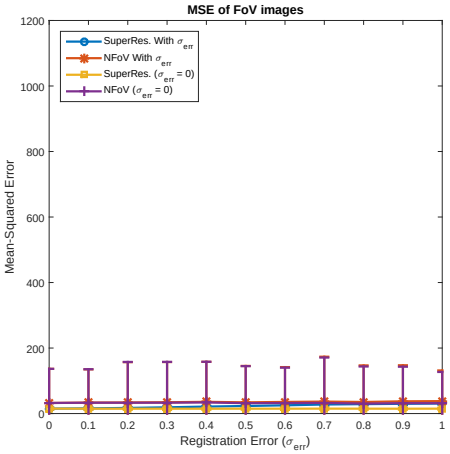


(j)

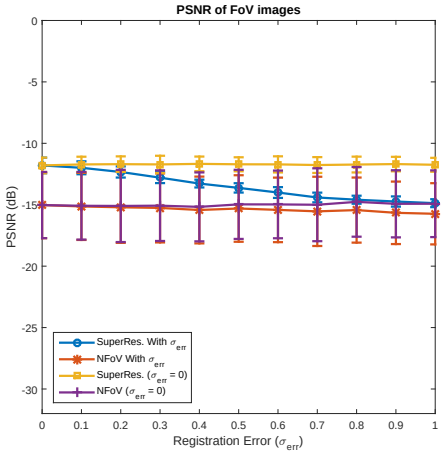


(k)

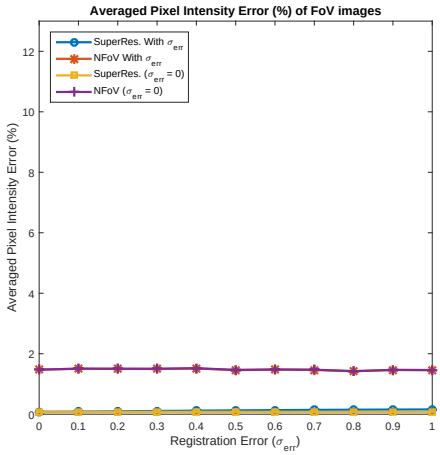
Figure C.23: Information quality information metric (mutual) JPEG (IQIMM JPEG) against varying realisations, as the image registration error increases from $0px$ to $1px$ (Figure C.23(a) - C.23(k))


MSE of FoV images
Mean-Squared Error
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

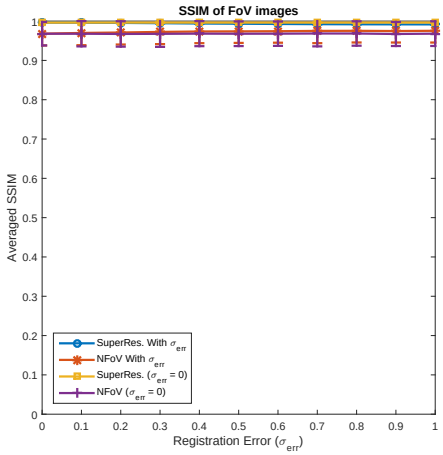
(a)


PSNR of FoV images
PSNR (dB)
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

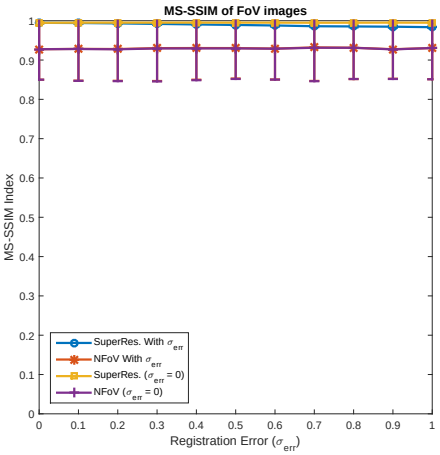
(b)


Averaged Pixel Intensity Error (%) of FoV images
Averaged Pixel Intensity Error (%)
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

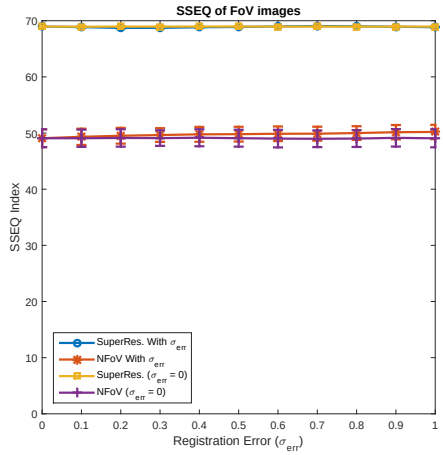
(c)


SSIM of FoV images
Averaged SSIM
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

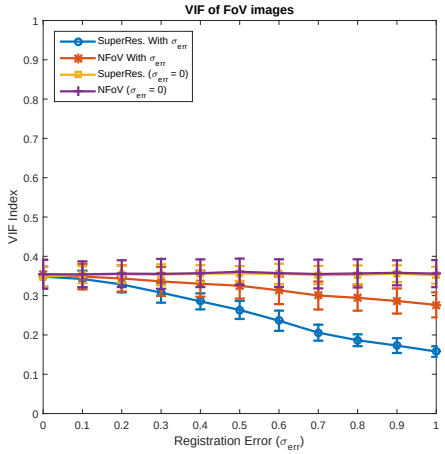
(d)


MS-SSIM of FoV images
MS-SSIM Index
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

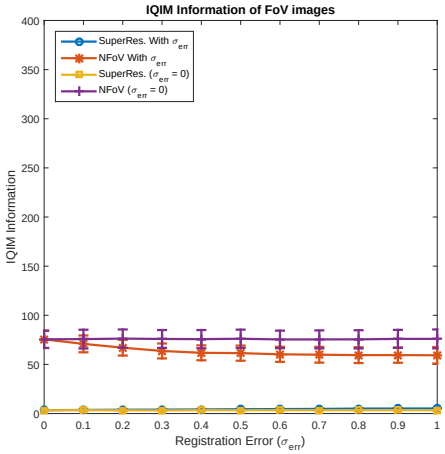
(e)


SSEQ of FoV images
SSEQ Index
Registration Error (σ_{err})
SuperRes. With σ_{err}
NFoV With σ_{err}
SuperRes. (σ_{err} = 0)
NFoV (σ_{err} = 0)

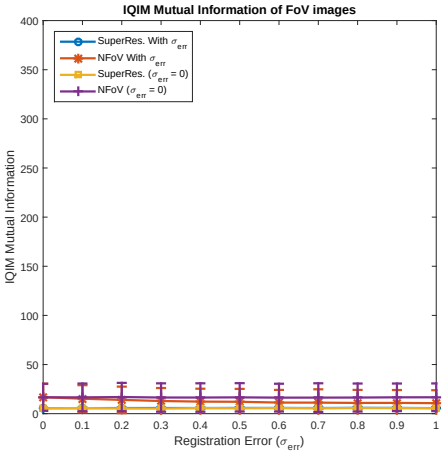
(f)



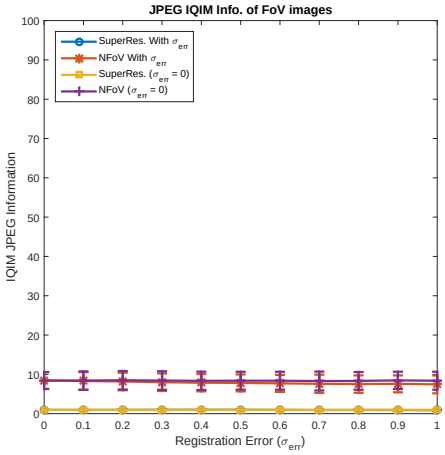
(g)



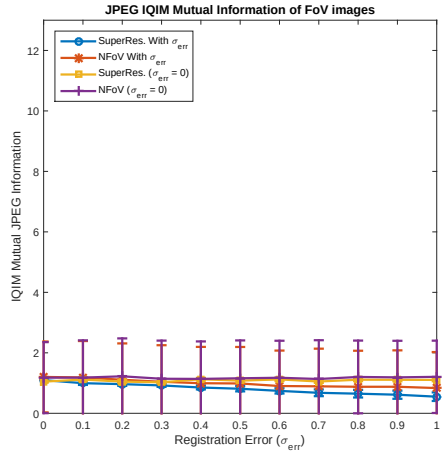
(h)



(i)



(j)



(k)

Figure C.24: Figures C.24(a) - C.24(k) showing the different image quality metrics used with error bars for the BGO_000[nfov] image. Each sub-figure contains the data for the perfect registration case, and the case with a registration error of $0.5px$, for the TA and SR techniques

Appendix D

Separated Images

(a) Original NFoV Chicago image

(b) Recovered NFoV Chicago image using TA with perfect registration

(c) NFoV Chicago image using SR with perfect registration

(d) Recovered NFoV Chicago image using TA with registration errors

(e) NFoV Chicago image using SR with registration errors

Figure D.1: Visual comparison of temporal averaging technique and super-resolution technique

(a) Original NFoV Croatia image

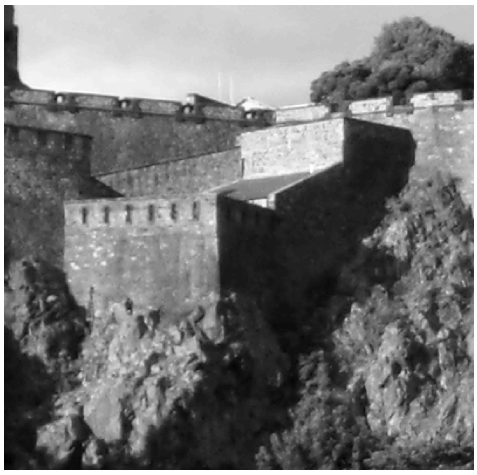
(b) Recovered NFoV Croatia image using TA with perfect registration

(c) NFoV Croatia image using SR with perfect registration

(d) Recovered NFoV Croatia image using TA with registration errors

(e) NFoV Croatia image using SR with registration errors

Figure D.2: Visual comparison of temporal averaging technique and super-resolution technique



(a) Original NFoV Edinburgh1 image

(b) Recovered NFoV Edinburgh1 image using TA with perfect registration

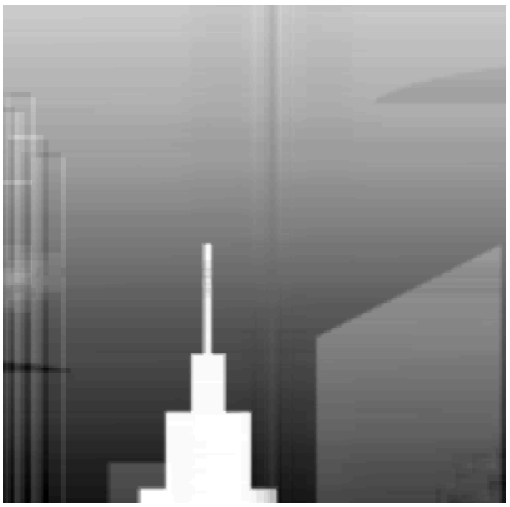
(c) NFoV Edinburgh1 image using SR with perfect registration

(d) Recovered NFoV Edinburgh1 image using TA with registration errors

(e) NFoV Edinburgh1 image using SR with registration errors

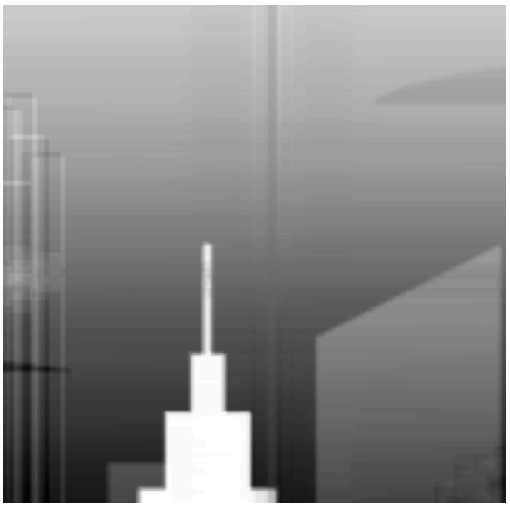
Figure D.3: Visual comparison of temporal averaging technique and super-resolution technique

(a) Original NFoV Night-cityscape image

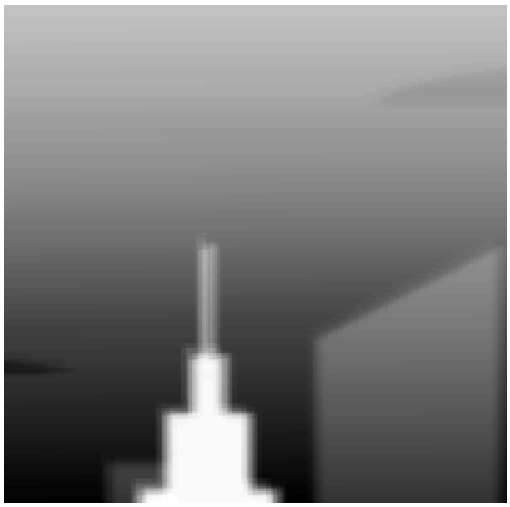


(b) Recovered NFoV Night-cityscape image using TA with perfect registration

(c) NFoV Night-cityscape image using SR with perfect registration



(d) Recovered NFoV Night-cityscape image using TA with registration errors



(e) NFoV Night-cityscape image using SR with registration errors

Figure D.4: Visual comparison of temporal averaging technique and super-resolution technique

(a) Original NFoV DSCN0026 image

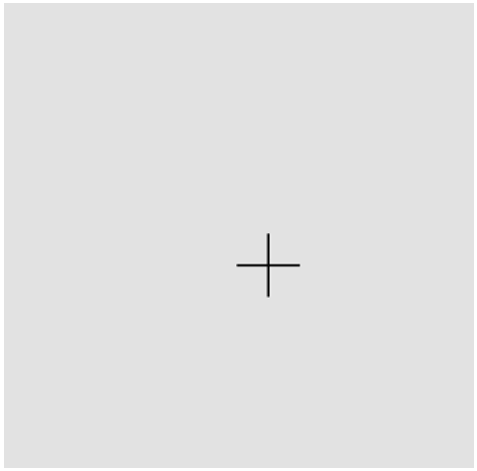
(b) Recovered NFoV DSCN0026 image using TA with perfect registration

(c) NFoV DSCN0026 image using SR with perfect registration

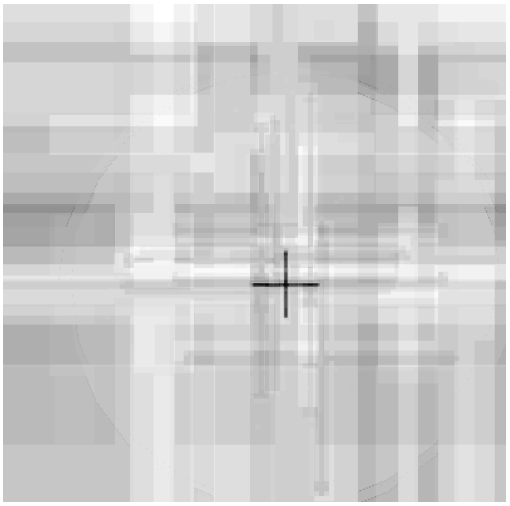
(d) Recovered NFoV DSCN0026 image using TA with registration errors

(e) NFoV DSCN0026 image using SR with registration errors

Figure D.5: Visual comparison of temporal averaging technique and super-resolution technique

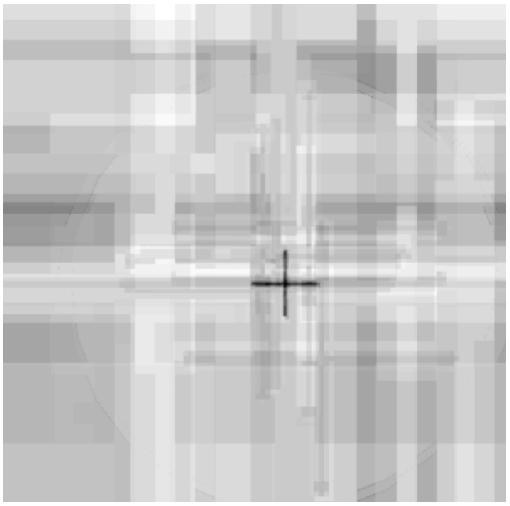


(a) Original NFoV BGO_000[nfov] image

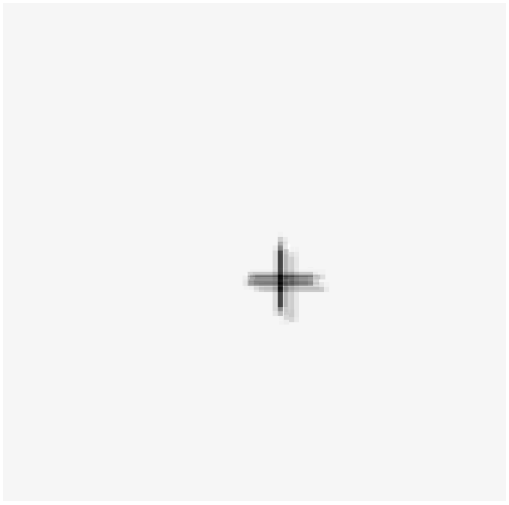


(b) Recovered NFoV BGO_000[nfov] image using TA with perfect registration

(c) NFoV BGO_000[nfov] image using SR with perfect registration



(d) Recovered NFoV BGO_000[nfov] image using TA with registration errors



(e) NFoV BGO_000[nfov] image using SR with registration errors

Figure D.6: Visual comparison of temporal averaging technique and super-resolution technique

Bibliography

[1] DigitalGlobe 2016 ©. Digitalglobe is the worlds trusted provider of high-resolution earth imagery, data, and analytics. `https://www.digitalglobe.com/`. Online; Accessed: 09-July-2016.

[2] ISO/IEC 2013. Information technology digital compression and coding of continuous-tone still images: Application to printing systems part 6. Technical report, ISO/IEC JTC 1/SC 29, Apr 2013. ISO/IEC 10918-6:2013.

[3] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1):90–93, Jan 1974.

[4] Fakhre Alam, Sami Ur Rahman, Shah Khusro, Sehat Ullah, and Adnan Khalil. Evaluation of medical image registration techniques based on nature and domain of the transformation. *Journal of Medical Imaging and Radiation Sciences*, 47(2):178 – 193, 2016.

[5] Y. Amit and A. Kong. Graphical templates for model registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):225–236, Mar 1996.

[6] A. Amruta, A. Gole, and Y. Karunakar. A systematic algorithm for 3-d reconstruction of mri based brain tumors using morphological operators and bicubic interpolation. In *Computer Technology and Development (ICCTD), 2010 2nd International Conference on*, pages 305–309, Nov 2010.

[7] J. Anderson. *Fundamentals of Aerodynamics*. McGraw-Hill Education, 2010.

[8] J.D. Anderson. *Aircraft performance and design*. McGraw-Hill international editions: Aerospace science/technology series. WCB/McGraw-Hill, 1999.

[9] Bernard Ans, Jeanny Hérault, and Christian Jutten. Adaptive neural architectures: detection of primitives. *Proceedings of COGNITIVA*, 85:593–597, 1985.

[10] Sebastian Anthony. DARPA shows off 1.8-gigapixel surveillance drone, can spot a terrorist from 20,000 feet. `http://www.extremetech.com/extreme/146909-darpa-shows-off-1-8-gigapixel-surveillance-drone-can-spot-a-terrorist-from-20000-feet`, Jan 2013. [Online, Accessed: 21-Mar-2016].

[11] army technology. Brimstone advanced anti-armour missile, united kingdom. http://www.army-technology.com/projects/brimstone/, 2016. [Online; Accessed: 16-Sept-2016].

[12] Donald G. Bailey and C. Bouganis. Tracking performance of a foveated vision system. In *4th International Conference on Autonomous Robots and Agents, ICARA 2009*, volume 1, pages 414–419. IEEE 2009, 2009.

[13] S.N. Balakrishnan, A. Tsourdos, and B.A. White. *Advances in Missile Guidance, Control, and Estimation*. Automation and Control Engineering. CRC Press, 2016.

[14] David Balihar. What is a pinhole camera? [pinhole.cz]. http://www.pinhole.cz/en/pinholecameras/whatis.html, 2015. 2016-6-30.

[15] D. I. Barnea and H. F. Silverman. A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, C-21(2):179–186, Feb 1972.

[16] Harry G Barrow, Jay M Tenenbaum, Robert C Bolles, and Helen C Wolf. Parametric correspondence and chamfer matching: Two new techniques for image matching. Technical report, DTIC Document, 1977.

[17] Adrien Bartoli, Mathieu Coquerelle, and Peter Sturm. *A Framework for Pencil-of-Points Structure-from-Motion*, pages 28–40. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[18] Guido Bartoli. Image registration techniques: A comprehensive survey. *Universit, degli Studi di Siena, Visual Information Processing and Protection Group*, 2007.

[19] K. G. Baum, M. Helguera, J. P. Hornak, J. P. Kerekes, E. D. Montag, M. Z. Unlu, D. H. Feiglin, and A. Krol. Techniques for fusion of multimodal images: Application to breast imaging. In *Image Processing, 2006 IEEE International Conference on*, pages 2521–2524, Oct 2006.

[20] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, June 2008. Similarity Matching in Computer Vision and Multimedia.

[21] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I*, pages 404–417, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[22] P. R. Beaudet. Rotationally invariant image operators. In *International Conference on Pattern Recognition (ICPR)*, pages 579–583, 1978.

[23] Doug Beizer, The Washington Post Company. BAE to develop surveillance system. http://www.washingtonpost.com/wp-dyn/content/article/2007/11/11/AR2007111101348.html, Nov 2007. [Online; posted 12-November-2007].

[24] Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.

[25] R. Berthilsson. Affine correlation. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, volume 2, pages 1458–1460, Aug 1998.

[26] Sombir Singh Bisht, Bhumika Gupta, and Parvez Rahi. Image registration concept and techniques: A review. *International Journal of Engineering Research and Applications*, 4(4):30–35, Apr 2014.

[27] J.H. Blakelock. *Automatic control of aircraft and missiles*. Wiley, 1965.

[28] Gunilla Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 10(6):849–865, 1988.

[29] Sean Borman and Robert L Stevenson. Super-resolution from image sequences-a review. In *Midwest Symposium on Circuits and Systems*, pages 374–378. Notre Dame, Indiana, 1998.

[30] D.L. Boslaugh. *When Computers Went to Sea: The Digitization of the United States Navy*. Perspectives Series. Wiley, 2003.

[31] A.C. Bovik, C.W. Chen, and D. Goldgof. *Advances in Image Processing and Understanding: A Festschrift for Thomas S Huang*. Series in Machine Perception and Artificial Intelligence. World Scientific, 2002.

[32] Robert A. Braeunig. Rocket & space technology, rocket propulsion. http://www.braeunig.us/space/propuls.htm, 2012. [Online; Accessed: 23-March-2016].

[33] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi. Sparse ica for blind separation of transmitted and reflected images. *Int J Imaging Syst Technol*, 15:84–91, 2005.

[34] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376, 1992.

[35] M. Brown and D. G. Lowe. Recognising panoramas. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1218–1225 vol.2, Oct 2003.

[36] Matthew Brown and David G Lowe. Invariant features from interest point groups. In *13th British Machine Vision Conference, BMVC 2002*, pages 253–262, Cardiff, UK, 2-5 September 2002.

[37] Grant Brunner. Check out google earths amazing new satellite imagery from landsat 8. `http://www.extremetech.com/extreme/230867-check-out-google-earths-amazing-new-satellite-imagery-from-landsat-8`, 28-June 2016. [Online; Accessed: 08-Feb-2016].

[38] G. Cai, B.M. Chen, and T.H. Lee. *Unmanned Rotorcraft Systems*. Advances in Industrial Control. Springer London, 2011.

[39] Aldo Camargo, Qiang He, and Kannappan Palaniappan. Performance evaluations for super-resolution mosaicing on uas surveillance videos. *International Journal of Advanced Robotic Systems*, 10(249), 2013.

[40] Aldo Camargo, Richard R Schultz, and Qiang He. Super-resolution mosaicking of unmanned aircraft system (uas) surveillance video using levenberg marquardt (lm) algorithm. In *Advances in Visual Computing, 6th International Symposium on Visual Computing, ISVC 2010*, pages 698–706, Las Vegas, NV, USA, Nov 2010. Springer.

[41] J.Y. Campbell, A.W.C. Lo, and A.C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

[42] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986.

[43] J. Cardoso. Source separation using higher order moments. In *International Conference on Acoustics, Speech and Signal Processing*, pages 2109–2112, 1989.

[44] J. F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Proceedings F - Radar and Signal Processing*, 140(6):362–370, Dec 1993.

[45] Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural computation*, 11(1):157–192, 1999.

[46] Ting Li Chang. *The IR missile (spin-scan and con-scan seekers) countermeasures*. PhD thesis, Monterey, California. Naval Postgraduate School, 1994.

[47] Ming-Jun Chen and Alan C Bovik. Fast structural similarity index algorithm. *Journal of Real-Time Image Processing*, 6(4):281–287, 2011.

[48] Yuan Chi. *Machine Learning Techniques for High Dimensional Data*. PhD thesis, The University of Liverpool, UK, Electrical Eng. & Computer Science dept., Oct 2015.

[49] Sunglok Choi, Taemin Kim, and Wonpil Yu. Performance evaluation of ransac family. *Journal of Computer Vision*, 24(3):271–300, 1997.

[50] Chin Seng Chua and Ray Jarvis. Point signatures: A new representation for 3d object recognition. *International Journal of Computer Vision*, 25(1):63–85, 1997.

[51] O. Chum and J. Matas. Matching with prosac - progressive sample consensus. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 220–226 vol. 1, June 2005.

[52] Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized ransac. In *Joint Pattern Recognition Symposium*, pages 236–243. Springer, 2003.

[53] Cisco. Understanding codecs: Complexity, hardware support, mos, and negotiation. `http://www.cisco.com/c/en/us/support/docs/voice/h323/14069-codec-complexity.html#mos`, February 2006. [Online; Accessed: 03/12/2015].

[54] Joint Photographic Experts Group Committee. Jpeg. `https://jpeg.org/index.html`. [Online; Accessed: 20-Apr-2016].

[55] P. Comon. Separation of stochastic processes. In *Workshop on Higher-Order Spectral Analysis*, pages 174–179, Vail, Colorado, 1989. IEEE.

[56] P. Comon. Independent component analysis, a new concept? In *Signal Processing*, volume 36, pages 287–314. Elsevier, 1994.

[57] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Elsevier Ltd, first edition, 2010.

[58] Pierre Comon. Independent component analysis. *Higher-Order Statistics*, pages 29–38, 1992.

[59] Curtiss-Wright. Size, weight and power (SWaP). `https://www.curtisswrightds.com/technologies/size-weight-and-power.html`, 2016. [Online; Accessed: 16-Sept-2016].

[60] Think Defence. Paveway vi. `http://www.thinkdefence.co.uk/uk-complex-weapons/brimstone/`, 2015. Online; Accessed:16-Sept-2016.

[61] Nathalie Delfosse and Philippe Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 45(1):59 – 83, 1995.

[62] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro. Principal curvature-based region detector for object recognition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[63] R. Deriche and T. Blaszka. Recovering and characterizing image features using an efficient model based approach. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on*, pages 530–535, Jun 1993.

[64] Konstantinos G Derpanis. Overview of the ransac algorithm. *Dept. of Electrical Eng., Lassonde School of Engineering*, 2010.

[65] John S DiMarco, Paul J Kemper Jr, and Lon N Pringle. Closed-loop guidance of imaging infrared missile seekers. In *AeroSense'99*, pages 254–265. International Society for Optics and Photonics, 1999.

[66] Neil A Dodgson. Image resampling. Technical Report UCAM-CL-TR-261, University of Cambridge, Computer Laboratory, 15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom, Aug. 1992.

[67] Silvano Donati. *Photodetectors: devices, circuits, and applications*. Prentice Hall PTR, 1999.

[68] Richard Dosselmann and Xue Dong Yang. A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, 5(1):81–91, 2009.

[69] C. S. Draper, W Wrigley, G Hoag, R. H. Battin, E. Miller, A. Koso, A. L. Hopkins, and W. E. Vander Velde. Apollo guidance and navigation. *Space Navigation Guidance and Control*, 1, June 1965.

[70] L. Dreschler and H.-H. Nagel. Volumetric model and 3d trajectory of a moving car derived from monocular tv frame sequences of a street scene. *Computer Graphics and Image Processing*, 20(3):199 – 228, 1982.

[71] M.R. Driels. *Weaponeering: Conventional Weapon System Effectiveness*. AIAA education series. American Institute of Aeronautics & Astronautics, 2013.

[72] Ronald G. Driggers, Carl E. Halford, Glenn D. Boreman, Douglas Lattman, and Kevin F. Williams. Parameters of spinning fm reticles. *Appl. Opt.*, 30(7):887–895, Mar 1991.

[73] C BROWN Duane. Close-range camera calibration. *Photogramm. Eng*, 37(8):855–866, 1971.

[74] Elan Dubrofsky. Homography estimation. Master's thesis, The University of British Columbia, Computer Science dept., 2009.

[75] Elan Dubrofsky and Robert J. Woodham. Combining line and point correspondences for homography estimation. In *Lecture Notes in Computer Science (LNCS)*, pages 202–213, Las Vegas, Nevada, USA, Dec 2008.

[76] Walter R. Dyer. Sample pages from modern missile analysis guidance, control, seekers, and technology. http://www.aticourses.com/sampler/Modern$_{\text{M}}$issile$_{\text{A}}$nalysis.pdf, 2004. [Online; Accessed: 09-Aug-2016].

[77] T. Dyson, W. Aslam, M. Aaronson, and R. Rauxloh. *Precision-Strike Technology and International Intervention: Strategic, Legal and Moral Implications*. Routledge Global Security Studies. Routledge, 2014.

[78] U. Engelke and H. J. Zepernick. Perceptual-based quality metrics for image and video services: A survey. In *Next Generation Internet Networks, 3rd EuroNGI Conference on*, pages 190–197, May 2007.

[79] UCL Dept. Medical Physics & Biomedical Engineering. Multimodality imaging. https://www.ucl.ac.uk/medphys/research/borl/imaging/multimodality. [Online, Accessed: 2016-03-17].

[80] A.M. Eskicioglu and P.S. Fisher. Image quality measures and their performance. *Communications, IEEE Transactions on*, 43(12):2959–2965, Dec 1995.

[81] R.M. Ezzeldeen, H.H. Ramadan, T.M. Nazmy, M. Adel Yehia, and M.S. Abdel-Wahab. Comparative study for image registration techniques of remote sensing images. *The Egyptian Journal of Remote Sensing and Space Science*, 13(1):31 – 36, 2010.

[82] H. Farid and E.H. Adelson. Separating reflections from images using independent components analysis. *Journal of the Optical Society of America*, 16:2136–2145, 1999.

[83] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, Oct 2004.

[84] Olivier Faugeras. *Three-dimensional Computer Vision: A Geometric Viewpoint*. Artificial intelligence. MIT Press, Cambridge, MA, USA, 1993.

[85] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[86] D. W. Fisher, R. F. Leftwich, and H. W. Yates. Survey of infrared trackers. *Appl. Opt.*, 5(4):507–515, Apr 1966.

[87] R. Fisher, S. Perkins, A. Walker, and E. Wolfart. Hypermedia image processing reference (hipr2). http://homepages.inf.ed.ac.uk/rbf/HIPR2/hipr$_{\text{t}}$op.htm, 2003. [Online; accessed 12-July-2016].

[88] E. L. Fleeman. Technologies for future precision strike missile systems. In *RTO Lecture Series 221*, NATO-RTO-EN-018. Georgia Institute of Technology, July 2001.

[89] E.L. Fleeman. *Missile Design and System Engineering*. AIAA education series. American Institute of Aeronautics and Astronautics, 2012.

[90] Leila MG Fonseca and BS Manjunath. Registration techniques for multisensor remotely sensed imagery. *PE & RS- Photogrammetric Engineering & Remote Sensing*, 62(9):1049–1056, 1996.

[91] Royal Air Force. Paveway vi. `http://www.raf.mod.uk/equipment/paveway4.cfm`, 2016. [Online; Accessed: 14-Sept-2016].

[92] Hassan Foroosh, Josiane B Zerubia, and Marc Berthod. Extension of phase correlation to subpixel registration. *IEEE transactions on image processing*, 11(3):188–200, 2002.

[93] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Always learning. Pearson Education, 2012.

[94] The MPEG Industry Forum. Understanding mpeg-4 | technologies, advantages and markets | an mpegif white paper. `https://www1.ethz.ch/replay/docs/whitepaper_mpegif.pdf`, 2005. [Online; Accessed: 05-Apr-2016].

[95] Steven E Fowler. Safety and arming device design principles. Technical report, DTIC Document, 1999.

[96] R.G. Frisius. *De radio astronomico et geometrico liber*. ap. Gul Cavellat, 1558.

[97] Jing Fu, Xiaojun Jing, Songlin Sun, Yueming Lu, and Ying Wang. C-surf: Colored speeded up robust features. In *International Conference on Trustworthy Computing and Services*, pages 203–210. Springer, 2012.

[98] Xinbo Gao, Wen Lu, Dacheng Tao, and Xuelong Li. Image quality assessment and human visual system. In *Visual Communications and Image Processing 2010*, pages 77440Z–77440Z. International Society for Optics and Photonics, 2010.

[99] Anna Geomi George and A Kethsy Prabavathy. A survey on different approaches used in image quality assessment. *International Journal of Emerging Technology and Advanced Engineering (IJETAE)*, 3(2), February 2013.

[100] Hassan Ghassemian. A review of remote sensing image fusion methods. *Information Fusion*, 32, Part A:75 – 89, 2016.

[101] Debabrata Ghosh, Naima Kaabouch, and William Semke. Super-resolution mosaicing of unmanned aircraft system (uas) surveillance video frames. *International Journal of Scientific & Engineering Research*, 4(2), Feb 2013.

[102] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 349–356. IEEE, 2009.

[103] ZENRIN EUROPE GmbH. Digital maps for japan, china, india & taiwan. `http://www.zenrin-europe.com/digital-maps.html`. Online; Accessed: 08-Sept-2016.

[104] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Pearson/Prentice Hall, 2008.

[105] A. Ardeshir Goshtasby. *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. Wiley-Interscience, 2005.

[106] A.A. Goshtasby. *Image Registration: Principles, Tools and Methods*. Advances in Computer Vision and Pattern Recognition. Springer London, 2012.

[107] Mohinder S. Grewal, Lawrence R. Weill, and Angus P. Andrews. *Global Positioning Systems, Inertial Navigation, and Integration*. John Wiley & Sons, Inc., 2nd edition, 2006.

[108] Elias J Griffith, Yuan Chi, Michael Jump, and Jason F Ralph. Equivalence of brisk descriptors for the registration of variable bit-depth aerial imagery. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 2587–2592. IEEE, 2013.

[109] Qualitative Reasoning Group. What is specific impulse? `http://www.qrg.northwestern.edu/projects/vss/docs/propulsion/3-what-is-specific-impulse.html`, 2016. [Online; Accessed: 08-Feb-2016].

[110] NORTHROP GRUMMAN. An/aaq-28(v) litening targeting pod. `http://www.northropgrumman.com/Capabilities/LITENING/Pages/default.aspx`. [Online; Accessed: 20-Sep-2013].

[111] J. Guehring. Reliable 3d surface acquisition, registration and validation using statistical error models. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 224–231, 2001.

[112] Joseph Gulick, W. Coleman Hyatt, and Jr. Oscar M. Martin. The talos guidance system. *John Hopkins APL Technical Digest*, 3(2):142–153, 1982.

[113] R. P. Hallion. Precision guided munitions and the new era of warfare. In *Air Power Studies Centre Working Papers*. Air Power Studies Centre, April 1993.

[114] N. Hanaizumi and S. Fujimur. An automated method for registration of satellite remote sensing images. In *Geoscience and Remote Sensing Symposium, 1993. IGARSS '93. Better Understanding of Earth Environment., International*, volume 3, pages 1348–1350, Aug 1993.

[115] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Proceedings of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[116] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[117] M Hassaballah, Aly Amin Abdelmgeid, and Hammam A Alshazly. Image features detection, description and matching. In *Image Feature Detectors and Descriptors*, pages 11–45. Springer, 2016.

[118] Phillip R. Hays. History of the talos missile guidance and homing systems. `http://www.okieboat.com/History%20guidance%20and%20homing.html`, 2009. [Online, Accessed: 02/03/2013].

[119] Qiang He and Richard Schultz. *Super-Resolution Reconstruction by Image Fusion and Application to Surveillance Videos Captured by Small Unmanned Aircraft Systems, Sensor Fusion and its Applications*. InTech, 2010.

[120] Jeff Hecht. *Understanding Lasers: An Entry-Level Guide*. John Wiley & Sons, Inc., 3rd edition, 2008.

[121] Jeanny Hérault, Christian Jutten, and Bernard Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *10 Colloque sur le traitement du signal et des images, FRA, 1985*. GRETSI, Groupe dEtudes du Traitement du Signal et des Images, 1985.

[122] Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in medicine and biology*, 46(3):R1, 2001.

[123] The Computer Vision Homepage. Computer vision test images. `https://www.cs.cmu.edu/~cil/v-images.html`. Online; Published: 30-June-2005.

[124] Y. C. Hsieh, D. M. McKeown, and F. P. Perlant. Performance evaluation of scene registration and stereo matching for cartographic feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):214–238, Feb 1992.

[125] Fang Rong Hu, Yi Xian Qian, Li Bao Yang, and Xiao Wei Cheng. Optimum design of laser encoding and decoding on laser beam-riding guidance. In *Frontiers of Manufacturing and Design Science II*, volume 121 of *Applied Mechanics and Materials*, pages 1028–1032. Trans Tech Publications, 1 2012.

[126] Andres Huertas, Mathias Bejanin, and Ramakant Nevatia. *Model Registration and Validation*, pages 33–42. Birkhäuser Basel, Basel, 1995.

[127] Vu Hung. Vu orthodontics - applied mechanics. `http://www.drhungvu.com/biomechanics`. [Online; Accessed: 18-Aug-2016].

[128] Q. Huynh-Thu and M. Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44(13):800–801, June 2008.

[129] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999.

[130] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430, 2000.

[131] Aapo Hyvrinen and Erkki Oja. Independent component analysis: A tutorial. `http://cis.legacy.ics.tkk.fi/aapo/papers/IJCNN99_tutorialweb/IJCNN99_tutorial3.html`, April 1999. [Online, Accessed: 07/02/2017].

[132] ImageProcessingPlace.com. Image databases. `http://www.imageprocessingplace.com/root_files_V3/image_databases.htm`. Online; Accessed: 26-Aug-2015.

[133] Apple Computer Inc. Mpeg-4 | the new standard for multimedia on the internet, powered by quicktime. `http://www.apple.com/quicktime/pdf/MPEG4_v3.pdf`, October 2003. [Online; Accessed: 05-Apr-2016].

[134] National Instruments. Peak signal-to-noise ratio as an image quality metric. Technical report, National Instruments, 11-Sept. 2013. [Online; Accessed: 13-Sep-2016].

[135] ECMA International. *JPEG File Interchange Format (JFIF)*. ECMA International, 1st edition, June 2009. `http://www.ecma-international.org/publications/techreports/E-TR-098.htm`.

[136] Michal Irani, Stephen Hsu, and P. Anandan. Mosaic-based video compression. In *Digital Video Compression: Algorithms and Technologies*, volume 2419, pages 242–253, 1995.

[137] Jyoti Joglekar and Shirish S Gedam. Area based image matching methods - a survey. *International Journal of Emerging Technology and Advanced Engineering*, 2(1), 2012.

[138] Alan D. Jones. Manual of photogrammetry, eds c.c. slama, c. theurer and s.w. hendrikson, american society of photogrammetry, falls church, va., 1980, fourth edition, 180 260mm, xvi and 1056 pages (with index), 72 tables, 866 figures. isbn 0 937294 01 2. *Cartography*, 12(4):258–258, 1982.

[139] Luo Juan and Oubong Gwun. A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4):143–152, 2009.

[140] Christian Jutten and Jeanny Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1 – 10, 1991.

[141] Vipin Kamble and K.M. Bhurchandi. No-reference image quality assessment algorithms: A survey. *Optik - International Journal for Light and Electron Optics*, 126:1090 – 1097, 2015.

[142] Syed Ali Khayam. The discrete cosine transform (dct): theory and application. *Michigan State University*, 114, 2003.

[143] Ryo Kurazume Ko, Ko Nishino, Zhengyou Zhang, and Katsushi Ikeuchi. Simultaneous 2d images and 3d geometric model registration for texture mapping utilizing reflectance attribute. In *In Proceedings of Fifth Asian Conference on Computer Vision*, pages 99–106, 2002.

[144] Jan J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984.

[145] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito. Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis. *IEE Proceedings I - Communications, Speech and Vision*, 140(1):19–24, Feb 1993.

[146] Carlo Kopp. Laser guidance. *Australian Aviation*, September 1981, 2005. [Online, Accessed: 02/03/2013].

[147] Carlo Kopp. Heat-seeking missile guidance. *Australian Aviation*, 1982(March), 1982. [Online; Accessed: 02-Mar-2013].

[148] Carlo Kopp. The sidewinder story; the evolution of the aim-9 missile. *Australian Aviation*, 1994(April), 1994.

[149] C. D. Kuglin and D. C. Hines. The phase correlation image alignment method. *Proc. Int. Conf. Cybernetics Society*, pages 163–165, 1975.

[150] K Divya Lakshmi and V Vaithiyanathan. Study of feature based image registration algorithms for navigation of unmanned aerial vehicles. *Indian Journal of Science and Technology*, 8(22), 2015.

[151] David C. Lay, Steven R. Lay, and Judi J. McDonald. *Linear Algebra and Its Applications*. Pearson, 5th ed. edition, 2015.

[152] Didier Le Gall. Mpeg: A video compression standard for multimedia applications. *Commun. ACM*, 34(4):46–58, April 1991.

[153] Mark Lee. Apple shares google china map partner in win for autonavi, 6th July 2012.

[154] Gerard Leng. DTS 5705 - MDTS Guidance, Aerodynamics and Control Website. `http://dynlab.mpe.nus.edu.sg/mpelsb/mdts/index.html`. [Online, Accessed: 2013-03-02].

[155] Sherry Van Liere. Contrast sensitivity. *Paraoptometric Resource Center*, 2010.

[156] Trimble Navigation Limited. Sketchup Pro 2015. `http://www.sketchup.com/`, 2015. [Online; Accessed: 10-Apr-2015].

[157] T. Lindeberg and J. O. Eklundh. Scale detection and region extraction from a scale-space primal sketch. In *Computer Vision, 1990. Proceedings, Third International Conference on*, pages 416–426, Dec 1990.

[158] Tony Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.

[159] Tony Lindeberg. *Scale-Space*. John Wiley & Sons, Inc., 2007.

[160] Tony Lindeberg. Scale selection. In Katsushi Ikeuchi, editor, *Computer Vision: A Reference Guide*, pages 701–713. Springer US, Boston, MA, 2014.

[161] Tony Lindeberg and Jonas Gårding. Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. *Image and vision computing*, 15(6):415–434, 1997.

[162] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, 29(8):856 – 863, 2014.

[163] Shizhong Liu and Alan C Bovik. Efficient dct-domain blind measurement and reduction of blocking artifacts. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(12):1139–1149, 2002.

[164] Y. Liu, H. Zhou, X. Su, M. Ni, and R. J. Lloyd. Transforming least squares to weighted least squares for accurate range image registration. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 232–239, June 2006.

[165] D. G. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150–1157 vol.2, 1999.

[166] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[167] D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[168] Jesus Malo, AM Pons, and JM Artigas. Subjective image fidelity metric based on bit allocation of the human visual system in the dct domain. *Image and Vision Computing*, 15(7):535–548, 1997.

[169] Nasim Mansurov. Understanding Aperture - A Beginner's Guide. `https://photographylife.com/what-is-aperture-in-photography`, Dec 2009. [Online; accessed 30-June-2016].

[170] Rafa K. Mantiuk, Anna Tomaszewska, and Radosaw Mantiuk. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum*, 31(8):2478–2491, Dec 2012.

[171] D. Marr. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. London, 2010.

[172] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences*, 207(1167):187–217, 1980.

[173] Prof David Marshall and Dr Kirill Sidorov. Cm3106 chapter 10: Discrete cosine transform. `https://www.cs.cf.ac.uk/Dave/Multimedia/PDF/10_DCT.pdf`. [Online; Accessed: 16-Apr-2016].

[174] Lockheed Martin. Lantirn er. `http://www.lockheedmartin.co.uk/us/products/LANTIRN.html`. [Online; Accessed: 20-Sep-2013].

[175] J Matas and O Chum. Randomized {RANSAC} with td,d test. *Image and Vision Computing*, 22(10):837 – 842, 2004. British Machine Vision Computing 2002.

[176] Jiri Matas, Ondrej Chum, Martin Urban, and Tomás Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10):761–767, 2004.

[177] MATLAB. *version 8.4.0 (R2014b)*. The MathWorks Inc., Natick, Massachusetts, 2014.

[178] MBDA. Subsystems & components. `http://www.mbda-systems.com/solutions-and-services/subsystems-components/`, 2016. [Online; Accessed: 12-Jun-2016].

[179] M. Mehta, E. J. Griffith, and J. F. Ralph. Geometric separation of superimposed images with varying fields-of-view. In *IEEE 17th International Conference on Information Fusion (FUSION)*, Salamanca, Spain, 2014. IEEE.

[180] M. M. Mehta, E. J. Griffith, S. Maskell, and J. F. Ralph. Geometric separation of superimposed images. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 1244–1251, July 2016.

[181] Nicolas Mellado, Dror Aiger, and Niloy J Mitra. Super 4pcs fast global pointcloud registration via smart indexing. *Computer Graphics Forum*, 33(5):205–215, 2014.

[182] Bill Messner and Dawn Tilbury. Aircraft Pitch: PID Controller Design. `http://ctms.engin.umich.edu/CTMS/index.php?example=AircraftPitch&section=ControlPID`, 2014. [Online; Accessed:01-Apr-2014].

[183] Bill Messner and Dawn Tilbury. Introduction: PID Controller Design. `http://ctms.engin.umich.edu/CTMS/index.php?example=Introduction&section=ControlPID`, 2014. [Online; Accessed: 10-Nov-2014].

[184] François G Meyer and Patrick Bouthemy. Region-based tracking using affine motion models in long image sequences. *CVGIP: Image understanding*, 60(2):119–140, 1994.

[185] K Mikolajczyk and C Schmid. An affine invariant interest point detector. In *Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 128–142. Springer-Verlag, 2002.

[186] Krystian Mikolajczyk. *Detection of local features invariant to affine transformations*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, France, 2002.

[187] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10):1615–1630, 2005.

[188] C. Mishra, M. Mehta, E. J. Griffith, and J. F. Ralph. Doing the right thing: Collision avoidance for autonomous air vehicles. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 2581–2586, Oct 2013.

[189] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[190] Jacqueline Le Moigne, Nathan S. Netanyahu, and Roger D. Eastman. *Image Registration for Remote Sensing*. Cambridge University Press, 2011. Cambridge Books Online.

[191] Ian Moir and Allan G. Seabridge. *Military Avionics Systems*. John Wiley & Sons, Ltd, 2006.

[192] A. K. Moorthy and A. C. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20(12):3350–3364, Dec 2011.

[193] Hans P. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. In *tech. report CMU-RI-TR-80-03*. Robotics Institute, Carnegie Mellon University & doctoral dissertation, Stanford University, September 1980.

[194] Scott M Murman, Michael J Aftosmis, and Marsha J Berger. Simulations of 6-dof motion with a cartesian method. *41st Aerospace Sciences Meeting and Exhibit*, 2003.

[195] D.R. Myatt, P.H.S. Torr, S.J. Nasuto, J.M. Bishop, and R. Craddock. Napsac: High noise, high dimensional robust estimation - it's in the bag. In *Proceedings of the British Machine Vision Conference*, pages 44.1–44.10. BMVA Press, 2002. doi:10.5244/C.16.44.

[196] Sudhakar Nair. *Mechanics of Aero-structures:*. Cambridge University Press, Jun 2015.

[197] NASA. Specific impulse. `https://www.grc.nasa.gov/WWW/K-12/airplane/specimp.html`, 2015. [Online; Accessed: 19-Oct-2015].

[198] Nikon. Nikon | Imaging Products | Coolpix L27. `http://imaging.nikon.com/lineup/coolpix/l/l27/`, 2014. [Online; Accessed: 01-Apr-2016].

[199] Dictionary of Military and Associated Terms. (2005). image degradation. `http://www.thefreedictionary.com/image+degradation`, 2005. [Online; Accessed: 13-Sep-2016].

[200] Hideo Ogawa. Labeled point pattern matching by fuzzy relaxation. *Pattern Recognition*, 17(5):569 – 573, 1984.

[201] Alan V Oppenheim and Ronald W Schafer. *Discrete-time signal processing*. Pearson Higher Education, 2010.

[202] Fletcher C. Paddison. The talos control system. *John Hopkins APL Technical Digest*, 3(2):154–156, 1982.

[203] Neil F Palumbo, Ross A Blauwkamp, and Justin M Lloyd. Basic principles of homing guidance. *Johns Hopkins APL Technical Digest*, 29(1):25–41, 2010.

[204] PM Panchal, SR Panchal, and SK Shah. A comparison of sift and surf. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(2):323–327, 2013.

[205] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.

[206] Andreas Parsch. Appendix 5: Guided bombs, paveway iii. `http://www.designation-systems.net/dusrm/app5/paveway-3.html`, 2006-2008. [Online; Accessed: 09-Aug-2016].

[207] Dinh Tuan Pham and P. Garat. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Transactions on Signal Processing*, 45(7):1712–1725, Jul 1997.

[208] N. Ponomarenko, V. Lukin, K. Egiazarian, Jaakko Astola, Marco Carli, and Federica Battisti. Color image database for evaluation of image quality metrics. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 403–408. IEEE, Oct 2008.

[209] Military Power. Aim-9 sidewinder. `http://militarypower.wikidot.com/aim-9-sidewinder`. [Online; Accessed: 03-Jun-2015].

[210] W. K. Pratt. Correlation techniques of image registration. *IEEE Transactions on Aerospace and Electronic Systems*, AES-10(3):353–358, May 1974.

[211] B Prescott and GF McLean. Line-based correction of radial lens distortion. *Graphical Models and Image Processing*, 59(1):39–47, 1997.

[212] Jason Ralph. Advanced system guidance lecture notes, 2014.

[213] Remya Ramachandran. A survey on image matching techniques. *International Journal For Research in Applied Science and Engineering Technology, IJRASET*, 2(5), May 2014.

[214] Sanjay Ranade and Azriel Rosenfeld. Point pattern matching by relaxation. *Pattern Recognition*, 12(4):269 – 275, 1980.

[215] K.R. Rao and P. Yip. *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic Press, 1990.

[216] B Srinivasa Reddy and Biswanath N Chatterji. An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE transactions on image processing*, 5(8):1266–1271, 1996.

[217] K Sreedhar Reddy and K Rama Linga Reddy. Enlargement of image based upon interpolation techniques. *International Journal of Advanced Research in Computer and Communication Engineering, IJARCCE*, 2(12), Dec. 2013.

[218] Judith Redi, Hantao Liu, Hani Alers, Rodolfo Zunino, and Ingrid Heynderickx. Comparing subjective image quality measurement methods for the creation of public databases. In *Image Quality and System Performance VII*, volume 7529, pages 01–11, 2010.

[219] Abdul Rehman and Zhou Wang. Reduced-reference image quality assessment by structural similarity estimation. *IEEE Transactions on Image Processing*, 21(8):3378–3389, 2012.

[220] E. Renner. *Pinhole Photography: Rediscovering a Historic Technique*. Photography techniques. Focal Press, 2000.

[221] E. Renner. *Pinhole Photography: From Historic Technique to Digital Application*. Taylor & Francis, 2012.

[222] Jr Roger L. Easton. Fundamentals of digital image processing. `https://www.cis.rit.edu/class/simg361/Notes_1222010.pdf`, 2010. [Online; accessed 10-July-2016].

[223] Karl Rohr. Localization properties of direct corner detectors. *Journal of Mathematical Imaging and Vision*, 4(2):139–150, 1994.

[224] A. Rosenfeld and A.C. Kak. *Digital Picture Processing*. Computer Science and Applied Mathematics. Elsevier Science, 2014.

[225] J. Roskam and C.T.E. Lan. *Airplane Aerodynamics and Performance*. Airplane design and analysis. Design, Analysis and Research Corporation, 1997.

[226] Peter J Rousseeuw. Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880, 1984.

[227] M. A. Saad, A. C. Bovik, and C. Charrier. A dct statistics-based blind image quality index. *IEEE Signal Processing Letters*, 17(6):583–586, June 2010.

[228] M. A. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Transactions on Image Processing*, 21(8):3339–3352, Aug 2012.

[229] Jonathan Sachs. Image resizing and resampling techniques. Technical report, Digital Light & Color, 2001.

[230] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of interest point detectors. *International Journal of computer vision*, 37(2):151–172, 2000.

[231] Robert A. Schowengerdt. {CHAPTER} 8 - image registration and fusion. In Robert A. Schowengerdt, editor, *Remote Sensing (Third edition)*, pages 355 – XXVI. Academic Press, Burlington, third edition edition, 2007.

[232] Jeff Scott. Aerospaceweb - missile guidance. `http://www.aerospaceweb.org/question/weapons/q0187.shtml`. [Published: 1-Aug-2004; Accessed: 18-Aug-2016].

[233] Eduard Serradell, Mustafa Özuysal, Vincent Lepetit, Pascal Fua, and Francesc Moreno-Noguer. Combining geometric and appearance priors for robust homography estimation. In *European Conference on Computer Vision*, pages 58–72. Springer, 2010.

[234] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[235] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, Nov 2006.

[236] Hamid R. Sheikh and Alan C. Bovik. a visual information fidelity approach to video quality assessment. In *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 23–25, 2005.

[237] Hamid Rahim Sheikh and Alan C Bovik. Image information and visual quality. *Image Processing, IEEE Transactions on*, 15(2):430–444, 2006.

[238] H.R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *Image Processing, IEEE Transactions on*, 14(12):2117–2128, Dec 2005.

[239] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik. Live image quality assessment database release 2. `http://live.ece.utexas.edu/research/quality`. [Online; Accessed: 18-Feb-2016].

[240] Y.Q. Shi and H. Sun. *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*. Image Processing Series. CRC Press, 1999.

[241] N.A. Shneydor. *Missile Guidance and Pursuit: Kinematics, Dynamics and Control*. Horwood engineering science series. Elsevier Science, 1998.

[242] A. Simper. Correcting general band-to-band misregistrations. In *Image Processing, 1996. Proceedings., International Conference on*, volume 2, pages 597–600, Sep 1996.

[243] Utkarsh Sinha. SIFT: Theory and Practice. `http://aishack.in/tutorials/sift-scale-invariant-feature-transform-introduction/`, 2010. [Online; accessed: 28-July-2016].

[244] George M. Siouris. *Missile Guidance and Control Systems*. Springer Science & Business Media, 2004.

[245] Stephen M. Smith and J. Michael Brady. Susan—a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.

[246] Jay N. Vizgaitis; Bjrn F. Andresen; Peter L. Marasco; Jasbinder S. Sanghera; Miguel P. Snyder, editor. *Foveated optics*, volume 9822, 2016.

[247] B.L. Stevens, F.L. Lewis, and E.N. Johnson. *Aircraft Control and Simulation: Dynamics, Controls Design, and Autonomous Systems*. Wiley, 2015.

[248] C. V. Stewart. Minpran: a new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):925–938, Oct 1995.

[249] Charles V Stewart. Robust parameter estimation in computer vision. *SIAM review*, 41(3):513–537, 1999.

[250] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends ® in Computer Graphics and Vision*, 2(1):1–104, 2006.

[251] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010. `http://szeliski.org/Book/`.

[252] Opencv Dev Team. Introduction to SURF (Speeded-Up Robust Features). `http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_surf_intro/py_surf_intro.html`, Nov 2014. [Online; Accessed:28-July-2016].

[253] Kian Kee Teoh, H. Ibrahim, and S. K. Bejo. Investigation on several basic interpolation methods for the use in remote sensing application. In *Innovative Technologies in Intelligent Systems and Industrial Applications, 2008. CITISIA 2008. IEEE Conference on*, pages 60–65, July 2008.

[254] Microsoft ©. Microsoft research cambridge object recognition image database. `https://www.microsoft.com/en-us/download/details.aspx?id=52644`. Online; Published: 17-May-2016.

[255] Microsoft ©. Recognition datasets. `https://www.cs.utexas.edu/~grauman/courses/spring2008/datasets.htm`. Online; Published: 30-Apr-2008.

[256] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla. Shape context and chamfer matching in cluttered scenes. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–127–I–133 vol.1, June 2003.

[257] J.-P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243 – 260, 1998.

[258] Kim-Han Thung and P. Raveendran. A survey of image quality measures. In *Technical Postgraduates (TECHPOS), 2009 International Conference for*, pages 1–4, Dec 2009.

[259] Philip HS Torr and Andrew Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[260] Tinne Tuytelaars and Krystian Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and trends ® in computer graphics and vision*, 3(3):177–280, 2008.

[261] Thales UK. Precision Strike, Lightweight Multi-Role Missile (LMM). `https://www.thalesgroup.com/sites/default/files/asset/document/LMMLand_05_12.pdf`. [Online; Accessed: 08-Feb-2014].

[262] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, Apr 1991.

[263] Prof. Kyros Kutulakos University of Toronto. Csc2503: Foundations of computer vision. `http://www.cs.toronto.edu/~kyros/courses/2503/Handouts/features.pdf`, 2015. [Online; Accessed: 21-Mar-2016.

[264] Defense Update. Future air-to-surface guided weapon (fasgw). http://defense-update.com/products/f/fasgw.html#more. [Online; Accessed: 09-Feb-2014].

[265] Christian J Van den Branden Lambrecht and Olivier Verscheure. Perceptual quality measure using a spatiotemporal model of the human visual system. In *Electronic Imaging: Science & Technology*, pages 450–461. International Society for Optics and Photonics, 1996.

[266] C.J. van den Branden Lambrecht. A working spatio-temporal model of the human visual system for image restoration and quality assessment applications. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 4, pages 2291–2294 vol. 4, May 1996.

[267] N. Vujovic and D. Brzakovic. Establishing the correspondence between control points in pairs of mammographic images. *IEEE Transactions on Image Processing*, 6(10):1388–1399, Oct 1997.

[268] Jirka Wagner. Weapon fact sheet 7: Paveway laser guided bombs. http://www.476vfightergroup.com/downloads.php?do=file&id=296&act=down&actionhash=guest. [Online; Accessed: 10-Sep-2013].

[269] Jirka Wagner. Paveway laser-guided bombs. http://www.military.cz/usa/air/in$_{s}$ervice/weapons/bombs/paveway/paveway$_{e}$n.htm, 2012. [Online; Accessed: 10-Sep-2013].

[270] Martin J Wainwright and Eero Simoncelli. Scale mixtures of graussians and the statistics of natural images. In *Neural information processing systems foundation*, 2000.

[271] J. Wang, J. F. Ralph, and J. Y. Goulermas. An analysis of a robust super resolution algorithm for infrared imaging. In *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*, pages 158–163, Sept 2009.

[272] Jianhua Wang, Fanhuai Shi, Jing Zhang, and Yuncai Liu. A new calibration model of camera lens distortion. *Pattern Recognition*, 41(2):607 – 615, 2008.

[273] Wen-Hao Wang and Yung-Chang Chen. Image registration by control points pairing using the invariant properties of line segments. *Pattern Recognition Letters*, 18(3):269 – 281, 1997.

[274] Yi Wang, R. Fevig, and R. R. Schultz. Super-resolution mosaicking of uav surveillance video. In *2008 15th IEEE International Conference on Image Processing*, pages 345–348, Oct 2008.

[275] Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? - a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26:98–117, Jan 2009.

[276] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. In *IEEE Transactions on Image Processing*, volume 13, pages 600–612, Apr 2004.

[277] Z. Wang, H. R. Sheikh, and A. C. Bovik. Objective video quality assessment. In *The Handbook of video databases: Design and Applications*, pages 1041–1078. CRC Press, Sept 2003.

[278] Zhou Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002.

[279] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Electronic Imaging 2005*, pages 149–159. International Society for Optics and Photonics, 2005.

[280] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. IEEE, 2003.

[281] Andrew B. Watson, James Hu, and John F Mcgowan Iii. Dvq: A digital video quality metric based on human vision. *Journal of Electronic Imaging*, 10:20–29, 2001.

[282] Neil J Weissman, Prem Soman, and Dipan J Shah. Multimodality imaging: opportunities and challenges. *JACC: Cardiovascular Imaging*, 6(9):1022–1023, 2013.

[283] G. J. Wen, J. j. Lv, and W. x. Yu. A high-performance feature-matching method for image registration by combining spatial and similarity information. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4):1266–1277, April 2008.

[284] Juyang Weng, Paul Cohen, Marc Herniou, et al. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on pattern analysis and machine intelligence*, 14(10):965–980, 1992.

[285] Matthew Weschler. How lasers work. `http://science.howstuffworks.com/laser.htm`, April 2000. [Online, Accessed: 02/03/2013].

[286] R. Westrum. *Sidewinder: Creative Missile Development at China Lake*. Naval Institute Press, 2013.

[287] Jr William Garten and Frank A. Dean. Evolution of the talos missile. *John Hopkins APL Technical Digest*, 3(2):117–122, 1982.

[288] A. Witkin. Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, volume 9, pages 150–153, Mar 1984.

[289] Andrew P. Witkin. Scale-space filtering. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'83, pages 1019–1022, San Francisco, CA, USA, 1983. Morgan Kaufmann Publishers Inc.

[290] Laurent Mazuray; Rolf Wartmann; Andrew P. Wood, editor. *Computational imaging: the improved and the impossible*, volume 9626, 2015.

[291] Medha V Wyawahare, Pradeep M Patil, Hemant K Abhyankar, et al. Image registration techniques: an overview. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2(3):11–28, 2009.

[292] Y. Xue and D. Morrell. Adaptive foveal sensor for target tracking. In *Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 848–852, Pacific Grove, CA, 2002.

[293] Cheng ye Wang, Hanfang Sun, Shiro Yada, and Azriel Rosenfeld. Some experiments in relaxation image matching using corner features. *Pattern Recognition*, 16(2):167 – 182, 1983.

[294] M. Young. Pinhole optics. *Appl. Opt.*, 10(12):2763–2767, Dec 1971.

[295] Matt Young. The pinhole camera: Imaging without lenses or mirrors. *The Physics Teacher*, 27(9):648–655, 1989.

[296] Paul Zarchan. *Tactical and Strategic Missile Guidance.* AIAA (American Institute of Aeronautics & Astronautics), 4th edition, 2003.

[297] W. Zhang and J. Kosecka. Generalized ransac framework for relaxed correspondence problems. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 854–860, June 2006.

[298] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[299] P.H. Zipfel. *Modeling and Simulation of Aerospace Vehicle Dynamics.* AIAA education series. American Institute of Aeronautics and Astronautics, 2007.

[300] Barbara Zitov and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977 – 1000, 2003.

[301] Marco Zuliani. Ransac for dummies. *Vision Research Lab, Dept. Electrical and Computer Engineering*, 2009.